

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:38:40 ; Search time 12 seconds
(without alignments)
1779.177 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLIRNGKIDTV.....INEFSQVTDWELNQGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	480.20	20.2	439	1	GLNA_THEMA
2	486	19.6	443	1	GLNA_CLOSA
3	464	19.5	446	1	GLNA_LACIA
4	443	18.6	446	1	GLNA_METVO
5	442.5	18.6	443	1	GLNA_BACSU
6	424.5	17.9	446	1	GLNA_STAM
7	423	17.8	446	1	GLNA_STAM
8	412	17.3	445	1	GLNA_LACDE
9	411.5	17.3	443	1	GLNA_BACCE
10	400.5	16.9	446	1	GLNA_METMP
11	398.5	16.8	454	1	GLNA_METJA
12	395	16.6	442	1	GLNA_METH
13	393.5	16.6	446	1	GLN2_MYCTU
14	391	16.5	443	1	GLNA_PYRKO
15	389.5	16.4	439	1	GLNA_PYRAB
16	389.5	16.4	439	1	GLNA_PYRFU
17	387.5	16.3	454	1	GLNA_HALVO
18	386.5	16.3	454	1	GLNA_HALNI
19	386	16.2	491	1	GLNA_ARCFU
20	381.5	16.1	439	1	GLNA_PYRHO
21	376	15.8	443	1	GLNA_PYRHO
22	353	14.9	435	1	GLN3_RHIME
23	341	14.4	472	1	YCJX_ECOLI
24	330.5	13.9	469	1	GLNA_AQUAE
25	330	13.9	435	1	GLN3_RHILP
26	327	13.8	471	1	GLNA_SULSO
27	312	13.1	469	1	GLN1_STRRP
28	310.5	13.1	473	1	GLNA_ANASP
29	308.5	13.0	474	1	GLN1_FRAAL
30	303	12.8	478	1	GLN1_MYCTU
31	302	12.7	470	1	GLNA_FREDI
32	301	12.7	469	1	GLN1_STRVR
33	300.5	12.6	473	1	GLNA_SULAC

RESULT 1
GLNA_THEMA
ID AC P36205;
DT 01-JUN-1994 (Rel. 29, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA OR TM0943.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RA MEDLINE=92287316; PubMed=1348781;
RX Sanangelantoni A.M., Forlani G., Ambroselli F., Cammarano P.,
RA Tiboni O.;
RT "The glnA gene of the extremely thermophilic eubacterium Thermotoga
RT maritima: cloning, primary structure, and expression in Escherichia
RT coli.";
RL J. Gen. Microbiol. 138:383-393(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=92287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton H.O., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -/- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -/- SUBUNIT: Oligomer of 12 subunits arranged in the form of two
CC hexagons (By similarity).
CC -/- SUBCELLULAR LOCATION: Cytoplasmic.
CC -/- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to a license@isb-sib.ch).
CC
CC EMBL; X60160; CAA42729.1; -;
CC EMBL; AE001758; AAD36024.1; -;
CC PIR; B72313; B72313.
CC HSP; P06201; ILGR.
CC TIGR; TM0943; -;
CC InterPro; IPR001691; GLN_synth.

ALIGNMENTS

34	299	12.6	469	1	GLNA_STRCO
35	293	12.3	473	1	GLNA_SYNP2
36	287.5	12.1	469	1	GLN1_RHIME
37	287	12.1	472	1	GLNA_HAEIN
38	283	11.9	468	1	GLNA_METCA
39	283	11.9	472	1	GLNA_PASMU
40	280	11.8	468	1	GLNA_AZOB
41	278.5	11.7	468	1	GLNA_VIBAL
42	276	11.6	467	1	GLNA_AZOV1
43	270	11.4	468	1	GLNA_AZOCA
44	269	11.3	481	1	GLNA_HELPY
45	267.5	11.3	469	1	GLNA_PROVU

P15106 streptomyc
P28605 synechococ
Q59747 rhizobium m
P43794 haemophilus
P15124 methylococ
Q9C1P2 pasteurella
P10583 azospirillu
P19204 vibrio algi
P22248 azotobacter
P94126 azorhizobiu
P94845 helicobacte
P28786 proteus vul

downscaled

```

DR InterPro: IPR004809; GlnA.
DR InterPro: IPR001637; GlnA adenyltn.
DR Pfam: PF00120; gln-synt; 1.
DR Pfam: PF03951; gln-synt; 1.
DR ProDom: PD001057; Gln_synt_C; 1.
DR TIGRFAMs: TIGR00653; GlnA; 1.
DR PROSITE: PS00180; GlnA; 1.
DR PROSITE: PS00181; GlnA_ATP; 1.
KW Ligase; Complete proteome.
FT CONFLICT 204 280 A -> G (IN REF. 1).
FT CONFLICT 280 280 A -> R (IN REF. 1).
FT CONFLICT 336 336 S -> T (IN REF. 1).
SQ SEQUENCE 439 AA; 50035 MW; ABE3E674BD2F2359 CRC64;

Query Match 20.2%; Score 480; DB 1; Length 439;
Best Local Similarity 29.9%; Pred. No. 1.1e-30;
Matches 138; Conservative 73; Mismatches 209; Indels 42; Gaps 9;

QY 3 ITVDLNNLRNGKIDTVVACVDMOGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIE- 61
DB 1 MTIETKRIIEENVFIRLOFTDINGTL-----KNLEITPDVLESWED 45

QY 62 GIAGGGEISS-VDTGYSDCHLCADLNSLHLLPWS-EGA-----VLAINPHNFTVSEPL 114
DB 46 GIMFDGSSIEGFVRIEESDMYKLPVLDTPAVLPMTVDGAKSARVICDVVTP---DGKPF 101
QY 115 FCSFVILWQOIERLANLKLKGLFASELEPNLENYKASQKHWNKLTQAQHHQWMI 174
DB 102 EGDPRYLRLRMKEAKQGLGYTPYAGPEMFFIIPINEKGEVPPEFLD-----HGGYFOL 155
QY 175 SASGGTETMRSVRNKLKLEAGILMEATHPEFLPSQHELNFVPADPLTMDRHHIAKHGVR 234
DB 156 LPLSKVEEIRDTAIALEKMGIIVEATHHEVAPSOHEVDFYDTFTLRTADNAQTQKLVK 215
QY 235 EMAEQSGWATFMAKLSSTALGNACHHMSLODAETKNAFYDQNDDEYGMSTLARNWIAG 294
DB 216 TMAIFGHYATFMKPFYGVNGSGMEVHMSL--FRGDKNAFYDPPDPLGLSKELRYFVG 273
QY 295 LLKYVEATYFFASYINSYKRLQPLTFAPTKCWAIDNRTSAFRLCNSKSEGINVELRG 354
DB 274 ILKHAKALAAVTNPTINSYKRLVPGYEPVYISWSVNSALRIPKARGKATRLRYRGP 333
QY 355 GADNLPYAFSAIIAGISGIEKLELPPASGNVY-----DKELPEFPNSLQNAH 407
DB 334 DPCSNLYLAPAAIILAGLDGINKIEPPAPVEENIYHMTSERREELNIESLPSLKEAVE 393
QY 408 LKESKMLNKTFGKILHYVNAANVEINEFSKQVTDWELNQ 449
DB 394 ELKKDDVIIDALGEHIFERFVEAAEKDWKEFSYVTINWELQR 435

RESULT 2
GLNA_CLOSA STANDARD; PRT; 443 AA.
AC P10656;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Clostridium saccharobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
NCBI_TaxID=169679;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=P262;
RX MEDLINE=88086901; PubMed=2891680;
RA Janssen P.J., Jones W.A., Jones D.T., Woods D.R.;
RT "Molecular analysis and regulation of the glnA gene of the Gram-
RL positive anaerobe Clostridium acetobutylicum.";
RL J. Bacteriol. 170:400-408 (1988).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +

L-glutamine.
-!- ENZYME REGULATION: THE ACTIVITY OF THIS ENZYME IS CONTROLLED
BY ADENYLATION. THE FULLY ADENYLATED ENZYME COMPLEX IS INACTIVE.
-!- SUBUNIT: OLIGOMER OF 12 SUBUNITS ARRANGED IN THE FORM OF TWO
HEXAGONS.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
-!- CAUTION: Was originally thought to originate from
C.acetobutylicum.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: M18966; AA23241.1; -.
PIR: A28676; AJCLOA.
HSP: P06201; ILGR.
InterPro: IPR001691; Gln_synth.
InterPro: IPR004809; GlnA.
Pfam: PF001637; GlnA adenyltn.
Pfam: PF03951; gln-synt; 1.
ProDom: PD001057; Gln_synt_C; 1.
TIGRFAMs: TIGR00653; GlnA; 1.
PROSITE: PS00180; GlnA; 1.
PROSITE: PS00181; GlnA_ATP; 1.
Ligase.
BINDING 392 392 AMP (UNDER CONDITIONS OF ABUNDANT
GLUTAMINE) (BY SIMILARITY).
SQ SEQUENCE 443 AA; 49682 MW; EDE5BF44495151FC CRC64;

Query Match 19.6%; Score 466; DB 1; Length 443;
Best Local Similarity 28.0%; Pred. No. 1.4e-29;
Matches 130; Conservative 84; Mismatches 211; Indels 40; Gaps 8;

QY 4 TYDELNNLRNGKIDTVVACVDMOGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGI 63
DB 5 TKEDIINLVKGVKGFIRLQFTDIFGLKNAITDKQL-----EKALDNECMFDGSSIDGF 60
QY 64 AGGGEIESVDGYSDCHLCADLNSLHLLPW--SEGAVLAISNPHNFTVSEPLFCSPVI 121
DB 61 -----VRTESDMNLRLNLDSPVFPWRPOQGVKARLICDVKPDGTFEGDPRHV 111
QY 122 LMQOIERLANLKLKGL-----FASELEPNLENYKASQKHWNKLTQAQHHQWMIISA 176
DB 112 L-----KANADAKELGYTMNVGPCEFFLE-----TDENGRATTNTQDKAGYFDLAP 160
QY 177 SSGIETFMRSVRNKLKLEAGILMEATHPEFLPSQHELNFVPADPLTMDRHHIAKHGVRBM 236
DB 161 TDLGENARRDMTALAEENGFEIEASHHEVAEQNEIDPKYGDALTADNIMTFKLWVKSI 220
QY 237 AEQSGWATFMAKLSSTALGNACHHMSLODAETKNAFYDQNDDEYGMSTLARNWIAGLL 296
DB 221 AQRHGLHASFPKPIFGINGSGMHVYNSL--FKQGNAPVDNDKNGLSKVAYQIAGLL 278
QY 297 KYVPEATYFFASYINSYKRLQPLTFAPTKCWAIDNRTSAFRLCNSKSEGINVELRIGGA 356
DB 279 KNIKGMAAVTNPTINSYKRLVPGYEPVYISWSVNSALRIPKARGKATRLRYRCPDP 338
QY 357 DLNYPYAFSAIIAGISGIEKLELPPASGNVY-----NDKELPEFPNSLQNAHLL 409
DB 339 SSNPYLVLACLQAGLDGINKNLQPPAEVEANIFAMTEQERKENGIDNLPNNLYEAVNYM 398
QY 410 KESKMLNKTFGKILHYVNAANVEINEFSKQVTDWELNQFNRY 454
DB 399 KENELAKKALGDHVYGVYVAGKAAEWDDYRTKVKHWELENYLNKY 443

```

```

RESULT 3
ID GLNA_LACLA STANDARD; PRT; 446 AA.
AC Q9CDL9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA OR L12200.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RX MEDLINE=21235186; PubMed=1137471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis sep. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- ENZYME REGULATION: DOES NOT SEEM TO BE REGULATED BY ADENYLATION.
CC -!- SUBUNIT: Oligomer of 12 subunits arranged in the form of two
CC hexagons (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006449; AAK06298.1; -
CC PIR; H86899; H86899.
CC HSSP; P06201; ILGR.
CC InterPro; IPR001691; Gln synth.
CC InterPro; IPR004809; GlnA.
CC Pfam; PF00120; gln-synt; 1.
CC ProDom; PD001057; Gln_synt_C; 1.
CC TIGRFAMs; TIGR00653; GlnA; 1.
CC PROSITE; PS00180; GlnA_1; 1.
CC PROSITE; PS00181; GlnA_ATP; 1.
CC Ligase; Complete proteome.
CC KW
CC SEQUENCE 446 AA; 49693 MW; COA953D5341AD5CF CRC64;

Query Match 19.5%; Score 464; DB 1; Length 446;
Best Local Similarity 23.0%; Pred. No. 2e-29;
Matches 137; Conservative 80; Mismatches 211; Indels 44; Gaps 11;

QY 1 MTITYDELNLRKGIQDVTVLACVDMQGRMLGKRL--TGRHFLGLDQKKISITFTVVAV 58
DB 1 MTITADIRRVKEDIKFLRLMTDILGLTKNVEVPATDQLKLENKMF-----GS 56
QY 59 TIEGIAGGVEISSVDGYSDCHLCADLSLHLLPWSE-----GAVLAISNPNHFTVSE 112
DB 57 SIEGF-----VRINESDMYLPDLDTWIVFPWGDYGVKAVICDVVTP-----GE 103
QY 113 PLFCSRVILMQOIERLANLKLKGL--FASELEFNLF--NETYKASQKHVKLTAOPHH 169
DB 104 PFAGDPGRVLKRNLSKMEKLGKFSNGLGPEPEFLFKLNENDEPTLEVNDKG----- 155
QY 170 QMNNISSAGSGETPWRSVRNKLEBAGILMEATHPEFLPSQHELFNVPADPLTMADRHIIA 229
DB 156 GYFDLAPDLAGNTRREIVNVLTDLGFVEVSHSHEVAGIQHEIDFKYANALKACDNQIQIF 215
QY 230 KHGVREMAEQSGMVATTFMAKLSTLALGNACHIHMSLQDAETEKNAFYDQNDYGMSTLAR 289

Db 216 KLVVKTITARKHGLHATFMAKPVHGINSGMHCHNSL--FTEDGANAFADPTGDMGLSDVAH 274
QY 290 NWIAGLLKYVPEATYFFASVINSYKRLQPLTFAPTKCCWALDNRTSAPRLCNSKSEGINV 349
DB 275 SFIAGLLKHAYFTAINPTVNSYKRLVPGVEADPVYVWAGRNRSPLIRVPASGLSTRV 334
QY 350 ELRIGGADLNPLYLAFSAIIAGISIGIBEKLELPPASGNVY--NDKE-----LPEFPNSL 402
DB 335 ELRAVDPTANPYLALAVLLAAGLDGVBEHLEAPEAIESNIYVMTTEERKAHGITDLPSTL 394
QY 403 QNATHLLKESKMLNKTEGKILHYHNAANVEINFEFSKQVTDWELNOGFNRY 454
DB 395 HNAVKALPEDIIVTEALGEHLVNFVFAKRIEWSAYAOQVFSOMEIDNLYLEY 446

RESULT 4
ID GLNA_METVO STANDARD; PRT; 446 AA.
AC P21154;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2188;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1537 / PS;
RX MEDLINE=90139872; PubMed=2575777;
RA Possot O., Sibold L., Aubert J.-P.;
RT "Nucleotide sequence and expression of the glutamine synthetase
RT structural gene, glnA, of the archaeobacterium Methanococcus voltae."
RL Res. Microbiol. 140:355-371(1989).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53509; CAA37585.1; -
CC PIR; A43995; A43995.
CC HSSP; P06201; ILGR.
CC InterPro; IPR001691; Gln synth.
CC InterPro; IPR004809; GlnA.
CC InterPro; IPR001637; GlnA adenyltn.
CC Pfam; PF00120; gln-synt; 1.
CC ProDom; PD001057; Gln_synt_N; 1.
CC TIGRFAMs; TIGR00653; GlnA; 1.
CC PROSITE; PS00180; GlnA_1; 1.
CC PROSITE; PS00181; GlnA_ATP; 1.
CC Ligase.
CC KW
CC SEQUENCE 446 AA; 50199 MW; 95449E4DE8542690 CRC64;

Query Match 18.6%; Score 443; DB 1; Length 446;
Best Local Similarity 28.7%; Pred. No. 9.6e-28;
Matches 142; Conservative 73; Mismatches 182; Indels 98; Gaps 17;

QY 1 MTITYDELNLR--IRNGKID-----TTVLACVDMQGRMLGKRLTGRHFLGLDQKK 48
DB 6 MALEYIKNNVRLRFQFVDIHGEKPNIAVPVKLTADGSEELMGVLENGLFFDG----- 60

```



```

RN  SEQUENCE FROM N.A.
RP  STRAIN=M50 / ATCC 700699, and N315;
RX  MEDLINE=21311952; PubMed=1418146;
RA  Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA  Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA  Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA  Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA  Sekizizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA  Kanelisa M., Yamashita A., Ohnima K., Furuya K., Yoshino C., Shiba T.,
RA  Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT  "Whole genome sequencing of methicillin-resistant Staphylococcus
RL  aureus";
RL  Lancet 357:1225-1240(2001).
CC  -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC  L-glutamine.
CC  -1- SUBUNIT: Oligomer of 12 subunits arranged in the form of two
CC  hexagons (by similarity).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF003361; BAB57472.1; -.
DR  EMBL; AF003133; BAB42404.1; -.
DR  PIR; H89905; H89905.
DR  HSSP; P06201; ILGR.
DR  InterPro; IPR001691; Gln_synth.
DR  InterPro; IPR004809; GlnA.
DR  InterPro; IPR001637; GlnA_adenyln.
DR  Pfam; PF00120; gln-synt; 1.
DR  Pfam; PF03951; gln-synt_N; 1.
DR  ProDom; PD001057; Gln_synt_C; 1.
DR  TIGRfam; TIGR00653; GlnA; 1.
DR  PROSITE; PS00180; GlnA_1; 1.
DR  PROSITE; PS00181; GlnA_ATP; 1.
DR  KW  Ligase; Complete proteome.
SQ  SEQUENCE 446 AA; 50854 MW; FF9F25631167903D CRC64;

Query Match 17.8%; Score 423; DB 1; Length 446;
Best Local Similarity 28.5%; Pred. No. 3.8e-26;
Matches 112; Conservative 82; Mismatches 169; Indels 30; Gaps 7;

QY 78 SDCHLCADLNSLHLLPWS--EGAVLAISNPHNFTSEPLFCSPRVILMQOIERLANLKLK 135
DB 68 SDMLHPDLDTWIFPWTACQGVKVARICDVYKDTGTPFGDPRANLKRVLKEMEDLQFT 127

QY 136 GL-FASELEFNLNFTYKSAQKHWKLNKLTAPQHHQ-----WNMISASSGIETFMRSVR 188
DB 128 DFNILGPEPEFFLKLDEK-----GEPTLELNDGGYFDLAPTDLGECRRDIV 175

QY 189 NKLEERAGILMEATHPELPQSHNELFVPADPLTWADRHIIAKGVREMAEQSGWATFWA 248
DB 176 LELEDMGFDIEAGHHEVAPQCHIDFKYADAVTACDNIQTQFLVVKTIARKHNLHATFMP 235

QY 249 KLSSTALGNACHTHMSLDQATEKNAFYDQNDYVGMSLTARNIAGLLKYVPEATYFFAS 308
DB 236 KPLFGVNGSGMHFNVSU--FKGENAFFDNTWGLTETAYQTAGVLKNQARGFTAVCPN 293

QY 309 YINSYKRLQPLTFAPTKCCWAIDNRISAFRLCNKSEGINVELRIGGADLNPLYAFSAII 368
DB 294 LVNSYKRLVGYEAPCVIANGSKNRPLIRVPSRGLSTRIEVRSDPAANPYMALAAAIL 353

QY 369 AAGISGIEEKLELPPASGVNY--NOKE-----LPEFPNSLQNAATHLLKESKMLNTFGE 421
DB 354 EAGLDGINKKLVPEPNQNIYEMNREEREAVGIQDLPTSLYLTALKAMRENEVIKALGN 413

QY 422 KLILHYVNAANVEINEFSKQVTDWELNQGFRY 454
DB 414 HIYNQFINSKSIWDYRRTQVSEWERDQYMKOY 446

RESULT 8
GLNA_LACDE STANDARD; PRT; 445 AA.
AC P45627;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Lactobacillus delbrueckii (subsp. bulgaricus).
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93073924; PubMed=1359838;
RA Ishino Y., Morgensthaler P., Hottinger H., Soell D.;
RT "Organization and nucleotide sequence of the glutamine synthetase
RT (glnA) gene from Lactobacillus delbrueckii subsp. bulgaricus.";
RL Appl. Environ. Microbiol. 58:3165-3169(1992).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- ENZYME REGULATION: DOES NOT SEEM TO BE REGULATED BY ADENYLATION.
CC -1- SUBUNIT: Oligomer of 12 subunits arranged in the form of two
CC hexagons (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; D10020; BAA00910.1; -.
DR  PIR; A48947; A48947.
DR  HSSP; P06201; ILGR.
DR  InterPro; IPR001691; Gln_synth.
DR  InterPro; IPR004809; GlnA.
DR  InterPro; IPR001637; GlnA_adenyln.
DR  Pfam; PF00120; gln-synt; 1.
DR  Pfam; PF03951; gln-synt_N; 1.
DR  ProDom; PD001057; Gln_synt_C; 1.
DR  TIGRfam; TIGR00653; GlnA; 1.
DR  PROSITE; PS00180; GlnA_1; 1.
DR  PROSITE; PS00181; GlnA_ATP; 1.
DR  KW  Ligase.
SQ  SEQUENCE 445 AA; 50133 MW; A57A9E11ABAF87E8 CRC64;

Query Match 17.3%; Score 412; DB 1; Length 445;
Best Local Similarity 28.3%; Pred. No. 2.8e-25;
Matches 133; Conservative 84; Mismatches 203; Indels 50; Gaps 16;

QY 3 ITYDELNNLIRNGKIDTVVLACVQMGRLMGKRLTGRHFLGLDQKKISISTFYAVTIEG 62
DB 5 ITEEIRKDVEEKVRFLELAFTDINGTL--KNL-----EVPVS-----QLDD 45

QY 63 IAGG--GVEISSVDYGY-----SDCHLCADLNSLHLLPWS---EGAV-LAINPHNFTVS 111
DB 46 VLGNQTRFDGSSID--GFVRLEESDMVLYPDLATLWLTAVTTVEEGTIGRLVCSVHN-VDG 103

QY 112 EPLFCSPRVILMQOIERLANLKLKGL-FASELEFNLNFTYKSAQKHWKLNKLTAPQHHQ 170
DB 104 TPFEGDPRNLKKVIAEMEGMFGSDFEIGFEAEFFLKE-----GKNGEETTQKSDHSS 157

QY 171 WNMISASSGIETFMRSVRNKLBEAGILMEATHPELPQSHNELFVPADPLTWADRHIIAK 230
```

```

Db 158 YFDWASEDEGAKRREIVETLEKLGFRVEAAHVGDCQOQIDFRFDNALATADKLOTFK 217
Qy 231 HGVREMAQSQGVATFMAKLSSTALGNACHIHMSLQDAETEKNAFYQNDDEYGMSTLARN 290
Db 218 MVKTIARKYHLHASFMAKPVVEGLANGMHTNMSL--LKDGKNAFYDKGQYNLSTALT 275
Qy 291 WIAGLLKVPVATYFFASYINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNKSEGINVE 350
Db 276 FLNGILEHARAITCVANPTVNSYKRLIPGFEAPYVISWASNRSPMVRIPNANEVGTGLE 335
Qy 351 LRIGGADINPLAFSAITAGISGIEE-KLELPPASGNVY--NDKELPE-----FPNSL 402
Db 336 MRSTDPANPYLLLSACUKAGTGKEGLKPMAPVTS--NLPEMTDDEKELGKLPBSTL 394
Qy 403 QNATHLLKESKMLNKTFCEKILHYVNAANVEINEFSKQVTDWELNOCQFN 452
Db 395 HNAIKAFKDEDEVKVSAFSEHIVDSFLELKEFETWALYTQSVSENEVKRYFN 444

RESULT 9
GLNA_BACCE STANDARD; PRT; 443 AA.
AC P19064;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
GN GLNA.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
RX MEDLINE=90034764; PubMed=2572584;
RA Nakano Y., Kato C., Tanaka E., Kimura K., Horikoshi K.;
RT "Nucleotide sequence of the glutamine synthetase gene (glnA) and its
upstream region from Bacillus cereus.";
RL J. Biochem. 106:209-215(1989).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
L-glutamine.
CC -1- ENZYME REGULATION: NOT REGULATED BY POST-TRANSLATIONAL
MODIFICATION AND NOT SUBJECT TO FEEDBACK INHIBITION.
CC -1- SUBUNIT: OLIGOMER OF 12 SUBUNITS ARRANGED IN THE FORM OF TWO
HEXAGONS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D00513; BAA00403.1; -
DR PIR; J00075; AJBSQU.
DR HSP; P06201; ILGR.
DR InterPro; IPR001691; Gln_synth.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR001637; GlnA_adenlytn.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; Gln_synt_C; 1.
DR TIGRFAMs; TIGR00653; GlnA; 1.
DR PROSITE; PS00180; GlnA; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
DR PROSITE; PS00182; GlnA_ADENYLATION; 1.
KW Ligase.
FT INIT_MET
SQ SEQUENCE 443 AA; 50064 MW; 14A9EAB948BE0657 CRC64;

```

```

Query Match 17.3%; Score 411.5; DB 1; Length 443;
Best Local Similarity 26.8%; Pred. No. 3.1e-25;
Matches 126; Conservative 82; Mismatches 212; Indels 51; Gaps 9;

Qy 4 TYDELANNLRNGKIDTVLACVDMQGRMGKRL-TGRHFLGLDOKKISISTFVAVTIEG 62
Db 4 TKEDIFLAKKEENVKYLQFTDILLGVIKNVEIPVSQLTKALDNKMMFDGS-----STEG 58
Qy 63 IAGGYEISSVDYTSYCHLCADLNSLHLLPWS--EGAVLAISNPHNFVTFSEPLFCSPRV 120
Db 59 F-----VRTEESDMYLPDLDTWVFPWTAEKVKVRLICDIYNADGTPPEGDPN 109
Qy 121 ILMOQIERLANLKLGL-FASELEFNLFNETYKSAQKHWKMLKTAQPHHOMNISASSG 179
Db 110 NLKRVLMKEMALGSDFNLGPEPEFFLF-----KVDEKGNPTLELNDNGG 154
Qy 180 I-----ETPMRSVRNKLKEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIAK 230
Db 155 YFDLAPMDLGENCRDIVLELEEMGFIEASHHEVAFQGHIEDPKYANAIKSCDDIQTFK 214
Qy 231 HGVREMAEQSQGVATFMAKLSSTALGNACHIHMSLQDAETEKNAFYQNDDEYGMSTLARN 290
Db 215 LVVKTIARKYHLHASFMAKPVVEGLANGMHTNMSL--FKNGENVFYQNGDLQSLDDARH 272
Qy 291 WIAGLLKVPVATYFFASYINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNKSEGINVE 350
Db 273 FIAGILKHAFAFTAVANPTVNSYKRLVPGVEAPCYVAWSAQNRSPLVRIIPASRGISTRVE 332
Qy 351 LRIGGADINPLAFSAITAGISGIEKLELPPASGNVY-----NDKELPEFPNSLQ 403
Db 333 VRSVDPAANPYLVNATLAAAGLGKIKNKLTPPAADVNRNIYVMTKEEREAGIVDLPATLA 392
Qy 404 NATHLLKESKMLNKTFCEKILHYVNAANVEINEFSKQVTDWELNOCQFNRY 454
Db 393 QALVTLOSNEVISNALGDHLEHFEAKFEPWFIDFTQVHQRDKQYMSLY 443

RESULT 10
GLNA_METMP STANDARD; PRT; 446 AA.
AC OS9648;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lu;
RX MEDLINE=99084965; PubMed=9864338;
RA Cohen-Kupiec R., Marx C.J., Leigh J.A.;
RT "Function and regulation of glnA in the methanogenic archaeon
Methanococcus maripaludis.";
RL J. Bacteriol. 181:256-261(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; AF062391; AAD04845.1; -

```


1990

Search completed: December 17, 2003, 22:43:22
Job time : 14 secs

```
RC STRAIN=GES / Orsay;
RX Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Priour D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.,
RT "An integrated analysis of the genome of the hyperthermophilic
RL archaeon Pyrococcus abyssi."
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ248288; CAB50513.1; --
DR PIR; C75009; C75009.
DR HSP; P06201; ILGR.
DR InterPro; IPR001691; Gln_synth.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR001637; GlnA_adenyltn.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; Gln_synt_C; 1.
DR TIGRFAMs; TIGR00653; GlnA; 1.
DR PROSITE; PS00180; GlnA_1; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
KW Ligase; Complete proteome.
FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT
FT GLUTAMINE) (BY SIMILARITY).
SQ SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;

Query Match 16.4%; Score 389.5; DB 1; Length 439;
Best Local Similarity 28.0%; Pred. No. 1.7e-23;
Matches 128; Conservative 73; Mismatches 195; Indels 61; Gaps 15;

Qy 20 VLACVDMQGRMGKRLTGRHFLGLDQKKISISTFVYAVTIEGIAGGGYEISSVDTGYSD 79
Db 17 VQLIFVDINGMPKGMETPASRL-----QEAIE-----DGISFDGSSVPGFQ-GIED 61

Qy 80 CHLC--ADLSHLPLPRSEGAVLAISNPHNV--TSEPLFCSPRVILMQQIERLANLKLK 135
Db 62 SDLIFKADPDPTVYVPPWDN-----VARVYGYIKDKRPGYADPRGLVKRVIEKLAEMGIK 116

Qy 136 GLFASELEFNLFNETYKSQKHWK-NLKTAQPHHQMWNISASSGIETFMRSVRNKLSEA 194
Db 117 AYIGPEPEFYLFK---KNGS---WELEIPDVGGYFDILTLDKADIK---REIAEYMPSP 167

Qy 195 GILMEATHPEFLPSQHELNFVPADPLTMADRHIIAKHGVRMAEQSGMVAFTFMKLSSTA 254
Db 168 GLVPEVLHHEVGKAQHEIDPRYDEALKTADNIIISFKYIVKAVAEVHGLYATFMKPIYGM 227

Qy 255 LGNACHTHMSLODAETEKNAFYDQNDYGMSTLARNWITAGLLKVPETATYFFASYINSYK 314
Db 228 PGNGMHLHLSLW--KEGENIF---KGEGLSETALHFITGGLLKHAKALTATNPTVNSYK 282

Qy 315 RLQPLTFAPTCKCWAIDNRTSAFLCNKSGINVELRIGGADINPVLAFSAIIAAGISG 374
Db 283 RLVGGEYAPVVISNGYKRSALIRVPAPWNGARIETRCPPDSPANPYFAFAILMAGLDG 342

Qy 375 IEKLELPPPPASGVNYNDKE-----LPEFPNSIQNATHLLKESKMLNKTFFGE----- 421
Db 343 IKHKVEPPAYVEENVYEMDECKRELGDITLPGSIGEALDELEKDKVVRREALGEAYKNFI 402

Qy 422 -----KLILHYNVAANVEINEFSKVQTDWELNOGF 451
Db 403 EYKRKEWESYLEYLEAKH--LPKDTKRVTEWELERYF 437
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 22:51:35 ; Search time 2955 Seconds
(without alignments)
3734.089 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MTITYDELNNLRNGKIDTV.....INFEKQVTDWELNQGPNRY 454

Scoring table:
BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Ygapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2_1/USPTO.spool/US10098602/runat_17122003_151237_28052/app.query.fasta_1.647
-DB=EST -QPMF=fastap -SUFFIX=rt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10098602 @CGN 1.1 2810 @runat_17122003_151237_28052 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_png:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	777	32.7	501	10	AW870989 ra53c11.y
2	696	29.3	442	9	AW828772 ra73h03.y
3	506.5	21.3	621	14	CB374375 ru99e08.y
4	430.5	18.1	626	14	CB279008 ru39e11.y
5	395	16.6	850	29	BZ701245 PUBM115TD
6	389	16.4	825	29	CC443184 PUBH069TD
7	370	15.6	851	29	CC381873 PUBFA22TD
8	347	14.6	815	29	BZ705035 PUBM084TD
9	330.5	13.9	690	12	BJ347175 BJ347175
10	328	13.8	699	29	BZ684738 PUBF560TD
11	321.5	13.5	1176	29	BZ569521 pac82-164
12	321	13.5	818	29	BZ687214 PUBD118TD
13	316.5	13.3	684	12	BJ358241 BJ358241
14	309.5	13.0	1138	10	BE366680 rockefell
15	295.5	12.4	584	9	AU039163 AU039163
16	293.5	12.4	883	29	CC439227 PUBM046TB
17	290.5	12.2	674	28	BH142677 TDGDR48TH
18	288	12.1	669	13	BU064742 Pgr_5_K19
19	285	12.0	682	29	BZ668122 PUBCR86TD
20	284	12.0	929	9	AA430929 M32 Where
21	279.5	11.8	705	14	CB970295 CAB10003
22	276.5	11.6	901	29	CC435159 PUBDF85TB
23	275.5	11.6	850	29	BZ704494 PUBM314TD
24	275.5	11.6	902	29	CC435165 PUBDF85TD
25	275.5	11.6	944	29	BZ679052 PUBG243TD
26	273	11.5	1124	29	BZ558114 pac81-60
27	271	11.4	676	10	BE431085 SUN011.B0
28	267.5	11.3	694	28	BH142971 TDGDU72TH
29	261	11.0	717	14	CB853860 UI-CP-DUI
30	260.5	11.0	727	9	AV928925 AV928925
31	251.5	10.6	818	12	BJ167896 BJ167896
32	245.5	10.3	679	10	BE918734 FM1_1_H12
33	242.5	10.2	576	12	BM348177 MEST287-E
34	242.5	10.2	910	29	CC381870 PUBFA22TB
35	235.5	9.9	794	11	AV109282 Zea mays
36	235	9.9	619	9	AI896662 EST266105
37	233.5	9.8	564	12	BM140402 WHE0483.e
38	233.5	9.8	581	12	BJ268011 BJ268011
39	231	9.7	572	13	BU833379 T047C06.P
40	230.5	9.7	659	13	BU061228 FgrN_5_K1
41	230.5	9.7	698	13	BQ281957 WHE3027.C
42	230.5	9.7	698	13	BQ282096 WHE3028.G
43	230.5	9.7	743	14	CD452699 WHE1119.H
44	230.5	9.7	749	13	BU490570 604131271
45	230	9.7	487	13	BU276033 Cr_Emb_03

ALIGNMENTS

RESULT 1
AW870989
LOCUS
DEFINITION
ra53c11.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
CDNA 5', similar to TR:088070.088070 PUTATIVE GLUTAMINE SYNTHETASE.
/, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AW870989
AW870989.1 GI:8005042
EST.
Meloidogyne incognita (southern root-knot nematode)
Meloidogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

```

REFERENCE
AUTHORS
1 (bases 1 to 501)
McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished
CONTACT: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Uma Rao and David Bird
(david.bird@ncsu.edu) at North Carolina State University. DNA
Sequencing by: Washington University Genome Sequencing Center St.
Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 410.
FEATURES
source
1..501
/organism="Meloiodogyne incognita"
/mol_type="mRNA"
/db_xref="taxon:6306"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLORL"
/clone_lib="Bird-Rao Meloiodogyne incognita J2"
/note="Vector: ZAP express - PBKCMV (Stratagene); Site 1:
EcoRI; Site 2: XhoI; Oligo (dT) primed library. cDNA was
contracted and cloned unidirectionally into the vector
within the 5' EcoRI and 3' XhoI sites. This library was
constructed by Dr. Uma Rao and Dr. David Bird at North
Carolina state University."
BASE COUNT
170 a 90 c 85 g 156 t
ORIGIN
Alignment Scores:
Pred. No.: 8.5e-79 Length: 501
Score: 777.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.70% Indels: 0
DB: 10 Gaps: 0
US-10-098-602A-2 (1-454) x AW870989 (1-501)
QY 307 AlaSerTyrIleAsnSerTyrIlyAsnArgLeuGlnProLeuThrPheAlaProThrIlyCys 326
Db 3 GCATCTTACATCACTCGTACAAAAGACTTCAACCGCTTACTTTTGGCCCAACAAATGT 62
QY 327 CysTTPAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerIlySerGluGly 346
Db 63 TGTTGGCAATTGACACCAACGACAGCGCTTTCGACTTTGTAATTCAAATCCAGGGA 122
QY 347 IleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrIleAlaPheSerAla 366
Db 123 ATTAATGTTGAGCTGCTGATTGGTGGCGTGTATTTGAACCCCTATTAGCTTTTCCGCA 182
QY 367 IleIleAlaAlaGlyIleSerGlyIleGluGlyLeuGluLeuProProAlaSer 386
Db 183 ATCATAGCTGCGAGGAATTAGCGGTATAGAAAGAAAGCTTGAACCTTCCCCCTCCGACT 242
QY 387 GlyAsnValTyrAsnAspIlyGluLeuProGluPheProAsnSerLeuGlnAsnAlaThr 406
Db 243 GGCATGTTTACATGATGAGGAATTAACCTGTAATTCCTTAATTCCTACAAATGCTACA 302
QY 407 HisLeuLeuIlyGluSerIlyMetLeuAsnIlyThrPheGlyGluIlyLeuIleLeuHis 426
Db 303 CATCTTCTAAAAGAAATCGAAATGCTGAATAAAACATTCGGGGAGAGTTGATTCTACAT 362

```

```

QY 427 TyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTropGlu 446
Db 363 TATGTAAACGCTGTAATGTTGAGATTAAATGAAATTTTCAAAACAAAGTTACTGACTGGAG 422
QY 447 LeuAsnGlnGlyPheAsnArgTyr 454
Db 423 CTTAATCAAGGATTATATAGATAT 446
RESULT 2
AW828772 442 bp mRNA linear EST 10-MAY-2001
LOCUS ra73h03.Y1 Bird-Rao Meloiodogyne incognita J2 Meloiodogyne incognita
DEFINITION cDNA 5' similar to TR:088070 O88070 PUTATIVE GLUTAMINE SYNTHETASE.
; mRNA sequence.
ACCESSION AW828772
VERSION AW828772.1 GI:7922569
KEYWORDS EST.
SOURCE Meloiodogyne incognita (southern root-knot nematode)
ORGANISM Meloiodogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloiodogyninae; Meloiodogyne.
REFERENCE 1 (bases 1 to 442)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished
CONTACT: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Uma Rao and David Bird
(david.bird@ncsu.edu) at North Carolina State University. DNA
Sequencing by: Washington University Genome Sequencing Center St.
Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
FEATURES
source
1..442
/organism="Meloiodogyne incognita"
/mol_type="mRNA"
/db_xref="taxon:6306"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLORL"
/clone_lib="Bird-Rao Meloiodogyne incognita J2"
/note="Vector: ZAP express - PBKCMV (Stratagene); Site 1:
EcoRI; Site 2: XhoI; Oligo (dT) primed library. cDNA was
contracted and cloned unidirectionally into the vector
within the 5' EcoRI and 3' XhoI sites. This library was
constructed by Dr. Uma Rao and Dr. David Bird at North
Carolina state University."
BASE COUNT 138 a 80 c 90 g 134 t
ORIGIN
Alignment Scores:
Pred. No.: 1.47e-69 Length: 442
Score: 696.00 Matches: 135
Percent Similarity: 93.20% Conservative: 2
Best Local Similarity: 91.84% Mismatches: 10
Query Match: 29.29% Indels: 0
DB: 9 Gaps: 0
US-10-098-602A-2 (1-454) x AW828772 (1-442)

```


Percent Similarity: 54.15% Conservative: 55
 Best Local Similarity: 34.30% Mismatches: 115
 Query Match: 16.62% Indels: 12
 DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x BZ701245 (1-850)

QY 183 PheMetArgSerValArgAsnLysLeuGluAlaGlyLeuMetGluAlaThrHis 202
 DB 849 TTCGCGCCGACGCGCATCGAGTGGATGCGATCTCGTGGTTCGCGAC 790

QY 203 ProGluPheLeuProSerGlnHisGluLeuAsnValProAlaAspProLeuThrMet 222
 DB 789 CACGAGGTGCGCGGTGACGAGATCGATCTCCGCTACGCGCGCGCTCGATG 730

QY 223 AlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMet 242
 DB 729 GCCGACAACTGATGCTTCGCTATCGTCAAGGAGTGGCGATCGGTGAGGTG 670

QY 243 ValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHis 262
 DB 669 TGGCGGAGTTCATGCCAAGCGCTCTCCGACCAACCGCGCTCGCGATGCACACGAC 610

QY 263 MetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheThrAlaLeuAspGluThr 282
 DB 609 ATGAGCTGTTCGAGGCGACAC-----AACGCTTCCACACCCCGCGACCGCATG 556

QY 283 GlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAla 302
 DB 555 CAGCTCTCGGACCGGACGAGTTCATCGCGGATCTCGACGACGCGCGGATC 496

QY 303 ThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAla 322
 DB 495 AGCCCGCTCACCAACGAGTGGTGAACAGTTACAGCGCTGATCCACGCGCGGAGGCA 436

QY 323 ProThrLysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeu----- 339
 DB 435 CCGACCGCGCGAGCTGGGTGGCGAAGCGCTCGCGCTCATCCGCTCGCGCTGAC 376

QY 340 CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsn 359
 DB 375 ACGCGAACAGGCGTCTCGGACGATCGAGTCCGACGCCGCGCTCGGTGTAAAC 316

QY 360 ProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeu 379
 DB 315 CCCTACCTGAGTTTCGGGTGCTCTCGCGCGGACTCAAGGCGCATCGCGGAGATAC 256

QY 380 GluLeuProProAlaSerGlyAsnValTyr-----AsnAsp 392
 DB 255 GAGTGTGCGGACGAGCGGAGGACGAGTCTGGGCGCTCACCTCGGCGGAGCTCGCACG 196

QY 393 LysGluLeuProGluPheProAsnSerLeuGlnAlaThrHisLeuLysGluSer 412
 DB 195 ATGGGTACACGAGTCTCGCGGAGCTCGCGAGCGCTCAAGGCGATGGAGTCTCG 136

QY 413 LysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsn 432
 DB 135 GAGTGTGCGGAGGACTCGGCGACGAGTGTTCGACTACTCTCTCCGCAACAGTGG 76

QY 433 ValGluLeuGlnPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
 DB 75 ACCGAGTGAACAGTACTACCGAGCGCGGTCAACCGGTTCGAGTCTCGGTGCG 25

RESULT 6
 CC443184/c
 LOCUS
 DEFINITION
 PUH069TD ZM 0.6-1.0 KB Zea mays genomic clone ZMMBTa453K17,
 genomic survey sequence.
 CC443184
 VERSION
 CC443184.1 GI:30946996
 GSS.
 KEYWORDS
 Zea mays
 Zea mays
 ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 825)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
 Maize Genomics Consortium
 Unpublished
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..825
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTa453K17"
 /clone_lib="ZM_0.6-1.0 KB"
 /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"
 BASE COUNT 121 a 258 c 298 g 148 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8,83e-34 Length: 825
 Score: 389.00 Matches: 93
 Percent Similarity: 54.65% Conservative: 54
 Best Local Similarity: 34.57% Mismatches: 110
 Query Match: 16.37% Indels: 12
 DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x CC443184 (1-825)

QY 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
 DB 810 CTGAGTGTGATGGGATCTCCGAGATTCCTGCACACCGAGGGTCCCGGTGACGAG 751

QY 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230
 DB 750 GAGATCGATCTCGCTACGCGGAGCGCTGTCTGATGGCCGACACGCTATGACCTTCGCG 691

QY 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
 DB 690 TACATCGTCAAGGAGTCCGATCGGTGAGGTGTGTGGCGAGCTTCATGCCCCAAGCCG 631

QY 251 SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
 DB 630 CTCCTCGACACACCGCGCTCGGCGATGCACACATGAGCTGTTCGAGCGGACACC 571

QY 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
 DB 570 -----AACGCTTCCACACCCCGACCGATCGAGCTCTCGGACACCGGCAACGAG 517

QY 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
 DB 516 TTCATCGCGGATCTCGACGACGCGCGGAGATCAGCGCGCTCACCACCAACGAGTGGGTG 457

QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrTrpAlaIle 330
 DB 456 AACAGTTTACAGCGCTGATCCACGCGCGGCGAGGACCGACCGCGCGGTTGGC 397

QY 331 AspAsnArgThrSerAlaPheArgLeu-----CysAsnSerLysSerGluGlyIle 347
 DB 396 GCGAACCGTCTCCGCTCATCCGCTCCCGCTGTACACGCCGGAACAGGCGTGTGTCGCGA 337

QY 348 AsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIle 367
 DB 348 AsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIle 367

Db 336 CGCATCGAGTCCGCGAGCCCGGACTCGGCTGTGTAACCCCTACTGACGTTCCGGTGCTG 277
 Qy 368 lleAlaAlaGlyleSerGlyleGluGlyLeuLeuGluLeuProProAlaSerGly 387
 Db 276 CTCGCGCGGAGTCAAGGCGATCCCGAGAAGTACGAGCTCCCGACGAGGCCGAGGAC 217
 Qy 388 AsnValTyr-----AsnAspLysGluLeuProGluPheProAsn 400
 Db 216 GACGCTGGGCGCTCACCTCGCGCGAGCGTCCGACGAGGCTACACCGAGTCCCGCGC 157
 Qy 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420
 Db 156 AGCCTCGCGAGCGCTCAAGCGGATGCTCGGAGCTCTCGGAGCTGCTCGCGAGCACTCGCG 97
 Qy 421 GluLysLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuAsnGluPheSerLys 440
 Db 96 GAGCAGCTGTTGCTGACTTCTCTCCGCAACAAAGTGACCGAGTGGAGCACTACCGCAGC 37
 Qy 441 GlnValThrAspTyrGluLeuAsnGln 449
 Db 36 CGGTCACCCCGTTCGAGCTCGTTCGC 10

RESULT 7

CC381873 851 bp DNA linear GSS 19-MAY-2003
 LOCUS PUHFA222TD ZM 0.6.1.0 KB Zea mays genomic clone ZMMBTA436D19,
 DEFINITION genomic survey sequence.

ACCESSION CC381873

VERSION CC381873.1 GI:30859448

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 851)
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
 A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

TITLE

Maize Genomics Consortium

JOURNAL

Unpublished

COMMENT

Other_GSSs: PUHFA222TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

1..851

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBTA436D19"

/clone_lib="ZM 0.6.1.0 KB"

/note="Vector: pCR4-toPO; Site_1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

121 a 269 c 308 g 153 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1,43e-31 Length: 851

Score: 370.00 Matches: 89

Percent Similarity: 54.30% Conservative: 50

Best Local Similarity: 34.7% Mismatches: 105

Query Match: 15.57% Indels: 12

DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x CC381873 (1-851)

Qy 183 PheMetArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHis 202

Db 799 TTCCGCCCCACCGCATCGACGCGTGGAGTCGATGGGCATCTCGTCGAGTTCTCGCAC 740
 Qy 203 ProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMet 222
 Db 739 CACGAGGGTGGCCCGGTACGACGAGATCTTCCTCCCTACGCCGCGCGCTGTCGATG 680
 Qy 223 AlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluInSerGlyMet 242
 Db 679 GCCGACACGTCATGACCTTCGCTACATCGTCAAGGAGGTCCGATCGGTGAGGGTGTG 620
 Qy 243 VallalThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHis 262
 Db 619 TGGCGCAGCTTCATGCCCAAGCGCTCTCCGACCAACGCGCGTCCGCGATGCACACGCAC 560
 Qy 263 MetSerLeuGlnAspAlaGluThrLysAsnAlaPheTyrAspGlnAsnAspGluTyr 282
 Db 559 ATGAGCCTGTTCCGAGGGCGACACC-----AACGCTTCCACACCCCGACGCCCGATG 506
 Qy 283 GlyMetSerThrLeuAlaArgAsnTyrIleAlaGlyLeuLeuLysTyrValProGluAla 302
 Db 505 CAGCTCTCGGACACCGGCAAGCAGTTTCATCGCGGCATCTTCGACGACGCGCGGAGATC 446
 Qy 303 ThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAla 322
 Db 445 AGCGCGCTCACCAACAGTGGGTGAACAGTTTACAAGCGCTGATCCACGCGCGCGGACCA 386
 Qy 323 ProThrLysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeu----- 339
 Db 385 CCGACCGCGCGAGCTGGGGTGGCGCAACCGCTCCGCGCTCATCCGCGCTCCCGCTGTAC 326
 Qy 340 CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsn 359
 Db 325 ACGCGACACAGGCGTCTCGCGCGCATCGAGTCCGAGCCCGCGACTCGCTGTGTAC 266
 Qy 360 ProTyrLeuAlaPheSerAlaIleAlaGlyIleSerGlyIleGluGluLysLeu 379
 Db 265 CCTACCTGAGCTTCCGCGTGTCTGCTCGCGCGGACTCAAGGGCATCCCGGAGAAGTAC 206
 Qy 380 GluLeuProProProAlaSerGlyAsnValTyr-----AsnAsp 392
 Db 205 GAGCTGCCCGCGAGGCGGAGACGCTCTGGGCGCTCAGCTCGGCGCGGCGTCCGACG 146
 Qy 393 LysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSer 412
 Db 145 ATGGGCTACCGGAGCTCCCGGCGAGCTCCCGGAGGCGCTCAAGGCGATGGAGTCTCG 86
 Qy 413 LysMetLeuAsnLysThrPheGlyGluLysLeuLeuHisTyrVal 428
 Db 85 GAGCTGGTCCGCGAGGCACTCGCGGAGCACGTTGTCGACTACTTCTC 38

RESULT 8

BZ705035/c

LOCUS BZ705035

DEFINITION

genomic survey sequence.

ACCESSION BZ705035

VERSION BZ705035.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 815)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick

A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

Maize Genomics Consortium

Unpublished

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..815
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMBPa089N23"
 /clone_lib="ZM 0.6 1.0 KB"
 /note="Vector: pCR1-TOPO; Site1: EcoRI; 0.6-1.0 kb high Cor selected genomic DNA library"
 126 a 256 c 293 g 140 t

BASE COUNT
 126 a 256 c 293 g 140 t

Alignment Scores:
 Pred. No.: 5,92e-29 Length: 815
 Score: 347.00 Matches: 83
 Percent Similarity: 54.73% Conservative: 50
 Best Local Similarity: 34.16% Mismatches: 98
 Query Match: 14.60% Indels: 12
 DB: 29 Gaps: 3

US-10-098-602A-2 (1-454) x BZ705035 (1-815)

QY	217	AlaAspProLeuThrMetAlaAspArgHisIleAlaAlaHisGlyValArgGluMet	236
DB	801	GCGAGCGCTGTCGATGCCGACACGTCATGACCTTCGCTACATCGTCAAGGAGTC	742
QY	237	AlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGly	256
DB	741	GCGATCGGTGAGGCTGTGGCGCAGCTTCATGCCCAAGCGCTCTCCGACACGCGCGC	682
QY	257	AsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrCluLysAsnAlaPheTyr	276
DB	681	TCGGCGATGCACACGACATGACCTGTTCGAGGCGGACACC-----AACGCCCTTCCAC	628
QY	277	AspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeu	296
DB	627	AACCCGACGACCCGATGAGCTCTCGGACCGGACGAGTTCATCCCGGATCTCTG	568
QY	297	LysTyrValProGluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeu	316
DB	567	CAGCAGCGCGCGAGATCAGCGCGCTCACCAACGATGGTGTAACAGTTACAAGCGCTG	508
QY	317	GlnProLeuThrPheAlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAla	336
DB	507	ATCCACGCGCGGAGGACCGACCGCGCGGACGTGGGGTGGCGGACCGCTCCGCGCTC	448
QY	337	PheArgLeu-----CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle	353
DB	447	ATCCGCTCCGCTGTACAGCGCGACAGAGCGTCTCGGACGATCGAGGTCCGACG	388
QY	354	GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSer	373
DB	387	CCGACTCGGCTTGTAAACCCCTACCTGACGTTTCGCGGTGCTGTCGCGCGGACTCAAG	328
QY	374	GlyIleGluLysLeuGluLeuProProAlaSerGlyAsnValTyr-----	390
DB	327	GGCATCCGCGAGAAGTACGAGCTGCCGACGAGCGGCGGACGACGCTCTGGCGCTCAC	268
QY	391	-----AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThr	406
DB	267	TCGGCGGAGCTCGCAGATGGGCTACACGAGCTCCCGGACGCTCCCGGAGGCGTCT	208
QY	407	HisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyCluLysLeuIleLeuHis	426
DB	207	AAGCGGATGAGTCTCGAGCTGGTTCGCGGACGCTCGCGGACGACGCTGTTCGACTAC	148
QY	427	TyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu	446

147 TTCCTCCGCAACAAGTGGACCGAGTGGAAACGACTACCGCAGCGGGTCAACCCCGTTCGAG 8
 447 LeuAsnGln 449
 87 CTCGGTGGC 79

RESULT 9
 BZ347175/c
 LOCUS
 DEFINITION
 Dictyostelium clone dda26g12 3', mRNA sequence.
 ACCESSION
 BZ347175
 VERSION
 BZ347175.1
 GI:19217682
 EST.
 SOURCE
 Dictyostelium discoideum
 ORGANISM
 Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE
 1 (bases 1 to 690)
 Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
 Full length cDNA of Dictyostelium discoideum at the aggregation stage
 JOURNAL
 Unpublished
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tehin@genes.nig.ac.jp.
 Location/Qualifiers
 1..690
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda26g12"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"
 BASE COUNT 204 a 110 c 112 g 262 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,59e-27 Length: 690
 Score: 330.50 Matches: 77
 Percent Similarity: 56.57% Conservative: 35
 Best Local Similarity: 38.89% Mismatches: 81
 Query Match: 13.91% Indels: 5
 DB: 12 Gaps: 4
 US-10-098-602A-2 (1-454) x BZ347175 (1-690)

QY	256	GlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPhe	277
DB	666	GGCTGCTCAGGACACATGTCATCAAAATTTCAATTTGTTTAAACCGGTAAGAACCTATTTC	601
QY	276	TyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeu	291
DB	606	CTCATGATGATCCGACCCCAATCATATGTCGACATTTTCAATCATTTGTTGTTGTTGTTCA	544
QY	296	LeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArg	311
DB	546	TTATTATTGTTGCTGATTTTACCATTTTTCGACCAACCAATCAATAGTATTATAACGT	481
QY	316	LeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSer	331
DB	486	TTAGTCGATGGGTATTGGGACCAACCAACCAATGCGGAATGGATATAGAACTGTA	421
QY	336	AlaPheArgLeuLysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGly	351
DB	426	GCATTAGAAATTTCAAGGT---GGTAAAGCAACAAGATCAGAAATTCAGAGTCACAGGA	371
QY	356	AlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGly	376

```

Db      369 TCAGATGTAATCTTACATTTCAATCGCGCTTCATTCGCTGCTGCTCTCTATGGTGC 310
Qy      376 GluGluLysLeuGluLeu---ProProAlaSerGlyAsnValTyrAsn-----Asp 392
Db      309 ATTAACAATTAGATTAAACAAACAAACCAATCATAGGTAAATAGTTATGATCTCTATAAA 250
Qy      393 LysGluLeuProGlu---PheProHsenSerLeuGlnAsnAlaThrHisLeuLeuLysGlu 411
Db      249 AAAGACTAGTTGAAGATTACAGATCATTCAGTGAATCAACTGAATTAATCATCAAAA 190
Qy      412 SerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuHisLysThrValAsnAlaAla 431
Db      189 TCAAGATTGCAAAAGAAATACCTTGGTGAAGAATTGTTGATCATTTGCTGAAACTGCT 130
Qy      432 AsnValGluLeuAsnGluPheSerLysGlnValThrAspTyrGluLeuAsnGln 449
Db      129 CGTTGGGAATATAGACAATTCATCAAGTTTCATAAATGGGAATTAGAAAGA 76

```

RESULT 10
BZ684738/c
LOCUS PUBF60TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta045123,
DEFINITION genomic survey sequence.

ACCESSION BZ684738
VERSION BZ684738.1 GI:28243411
KEYWORDS GSS.

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 699)

REFERENCE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and Bennetzen, J.

Maize Genomics Consortium

Unpublished

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5943

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

source

1..699

/organism="Zea mays"

/mol_type="genomic DNA"

/strains="B73"

/db_xref="taxon:4577"

/clone="ZMMBta045123"

/clone_lib="ZM.0.6 1.0 KB"

/notes="Vector: pCR4-toPO; Site 1: EcoRI; 0.6-1.0 kb high

Cot selected genomic DNA library"

98 a 218 c 252 g 131 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 7,116-27 Length: 699

Score: 328.00 Matches: 79

Percent Similarity: 54.31% Conservative: 47

Best Local Similarity: 34.05% Mismatches: 94

Query Match: 13.80% Indels: 12

DB: 29 Gaps: 3

US-10-098-602a-2 (1-454) x BZ684738 (1-699)

Qy 217 AlaAspProLeuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMet 236

Db 697 GCAGACGCGCTGCGAGTCCGACAAACGTCATCCCTCGCTACATCGTCAGGAGGTC 638

Qy 237 AlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGly 256

```

Db      637 GCGATCGGTGAGGTGTGTGGCGGACGTTTCATCCCAAGCGCGCTCTCGGACCCACGCGGC 578
Qy      257 AsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyr 276
Db      577 TCGGCGATGCACACGACATGAGCTGTTCGAGGGCGACACC-----AACCCCTTCAC 524
Qy      277 AspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTyrPileAlaGlyLeuLeu 296
Db      533 AACCCCGACGACCGGATCGAGCTTCGGACACCGGACAGCAGTTTCATCGCGGCATCTCG 464
Qy      297 LysTyrValProGluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeu 316
Db      463 CAGCACGCGCGGAGATCAGCGCGCTCACCACCAACGAGTGGGTGAACAGTTTACAAGGCGCTG 404
Qy      317 GlnProLeuThrPheAlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAla 336
Db      403 ATCCACGCGCGGAGGACCGACCGCGGAGTGGGTGGCGGCGGACCGCTCCCGGCTC 344
Qy      337 PheArgLeu-----CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle 353
Db      343 ATCCGCGCTCCGCTGTACACGCGGACCAAGGGGTGTCTCGGACGCGATCGAGGTCCGCGC 284
Qy      354 GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaGlyLysSer 373
Db      283 CCGGACTCGGCTGTAAACCCCTACCTGAGCTTGGCGGTCTGCTCGCGCGGAGCTCAAG 224
Qy      374 GlyIleGluGluLysLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db      223 GGCATCGCGGAGAGTACGAGCTGCCGACGAGCGGAGGAGCGTCTGGGCGCTCACC 164
Qy      391 -----AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThr 406
Db      163 TCGGCGGAGCGTCGCGACGATGGGTACACCGAGCTGCCCGGAGGCTCGCGGAGGCGCTC 104
Qy      407 HisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuHis 426
Db      103 AAGGCGATGGAGTCTCTCGGAGTGGTGGCGGAGCACTCGGCGGAGCAGCTGTTCGACTAC 44
Qy      427 TyrValAsnAlaAlaAsnValGluLeuAsnGluPhe 438
Db      43 TTCTCTCCGACACAGTGGACCGAGTGGACGACTAC 8

```

RESULT 11

BZ569521/c

LOCUS

DEFINITION

pacS2-164_871.s2 pacS2-164 Pseudomonas aeruginosa genomic clone

pacS2-164_871, genomic survey sequence.

ACCESSION BZ569521

VERSION BZ569521.1 GI:27204274

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1176)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

source

1..1176

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"


```

/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164.871"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."
BASE COUNT      221 a   334 c   366 g   255 t
ORIGIN
Alignment Scores:
Pred. No.:      9.22e-26      Length:      1176
Score:          321.50        Matches:      95
Percent Similarity: 46.52%    Conservative: 52
Best Local Similarity: 30.06%  Mismatches:  135
Query Match:    13.53%       Indels:      34
DB:            29           Gaps:        8

US-10-098-602A-2 (1-454) x BZ569521 (1-1176)
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 936 GAGGTTCGAAGTTCTCTACTTCAATCGGACTCAAGGGCA----- 898
Qy 161 AsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnLysSerAlaSerGlylle 180
Db 897 ACCCTTAACGGGGCGCTCGTATGCCGCTTG-----GTGCGGGCGATTTCTTGGG 844
Qy 181 GluThrPheMetArgSerValArg-----AsnLysLeuGlu 192
Db 843 AAAGGTTCCGCAAGTCGTCGAAGTTTATTCATCGGAGGACCTGGACGAATTACGT 784
Qy 193 -GluAla-----GlylleLeuMetGluAla 201
Db 783 CGAATGCTTCAGGACTTCATCGACGGCGCGGCCCGAGGCTTCGCGCGGAAGCAAT 724
Qy 201 rHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProLa---AspProLe 220
Db 723 CGTCGCCGAATCCGACCGCGGAAGTTTCGAGTCAACCTCAACACCGTCAACGACGCCCT 664
Qy 220 uThrMetAlaAspArgHisLlelleAlaLysHisGlyValArgGluMetAlaGluInSe 240
Db 663 CAAGGCTGCGACCAACCGGGTGCTGCTCAAGCGGCTGTCGAAGAACATCGCTACGACCA 604
Qy 240 rGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHi 260
Db 603 CGAGATGACACCACTTCATGGCCAGCCCTATCCGGCCGAGCGCGGGAACGACTGCA 544
Qy 260 sIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAs 280
Db 543 CGTGATATCTCGTGTCTGCAC---AAGCATGGCAACAACATCTTCCACGAGGATCC 487
Qy 280 pGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPr 300
Db 486 CGAG-----CAGAACCGCGCATTCGCCCATCGCATCGCGGTGTCTCGAGACCCGTCC 433
Qy 300 oGluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuTh 320
Db 432 GGCCTCCATGGCTTCTCTGCCGAGCTCACTCTACCGCGCTTCGGTTCGCAGTT 373
Qy 320 rPheAlaProThrLysCysTyrTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCy 340
Db 372 CTACGTCCGAACACCGCGAGCTGGGCGCTGGACAACCGCACCGCTGGCGCTCGCGTCC 313
Qy 340 sAsnSerLysSerGluGlylleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPr 360
Db 312 CACCGGACCGCCGAGCGGTGACCCCTGGAAATCATCGGGTGCCGCGGCGCCGCAACCC 253
Qy 360 oTyrLeuAlaPheSerAlaIlelleAlaAlaGlylleSerGlylleGluGluLysLeuGl 380
Db 252 CTACTCTACTGCGAGCGGTGCTGGCAGGGCTTCATCAGGGGTTCACCAACAGGTGCA 193
Qy 380 uLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAs 400

```

```

192 GCCGGCGCCCGCGATCGAAGCAACTCTCTAGACGACATGGAG---CCGAGCCTCGCCGAA 136
400 nSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGl 420
135 CAATTGCGGAGCCCTCGCGAGCTGGACGAAGCGAGATCATGCGGAGTACATCGA 76
420 yGluLysLeuLysHisTyrValAsnAlaAlaAsnValGluLle 435
75 CCGAAGTACATCGACATCTTCGTGCGCTGCAAGAAAGCGAGCTG 30

RESULT 12
BZ687214 818 bp DNA linear GSS 05-FEB-2003
LOCUS PUB118TD.ZM 0.6.1.0 KB Zea mays genomic clone ZMMB7a029D11,
DEFINITION genomic survey sequence.
ACCESSION BZ687214
VERSION BZ687214.1 GI:28248163
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 818)
AUTHORS (Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
source
1..818
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMB7a029D11"
/clone_lib="ZM 0.6.1.0 KB"
/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"
BASE COUNT 164 a 287 c 232 g 135 t
ORIGIN
Alignment Scores:
Pred. No.: 5.86e-26 Length: 818
Score: 321.00 Matches: 86
Percent Similarity: 49.25% Conservative: 45
Best Local Similarity: 32.33% Mismatches: 121
Query Match: 13.51% Indels: 14
DB: 29 Gaps: 3
US-10-098-602A-2 (1-454) x BZ687214 (1-818)
Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnIleGluArgLeuAlaAsnLeu 132
Db 36 CCCATCGAAGTGTGCGCGCAACGCTCTGAAGAAGTCTCTGAGTGTGTACACCGACG 95
133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLys 152
Db 96 GGTGCGCGGATCGTGGCGCGGAAATGAGTTTACCTG-----ACC 140
153 SerAlaSerGlnLysHisTyrLysAsnLysLysThr-----Ala 165
141 CACGGTGGCAACCGCGGACCTGCGCGTGAACCCACGATAGCCGCTCGGGCGCGCC 200
166 GlnProHisHisGlnTrpMetAsnLysSerAlaSerGlylleGluThrPheMetArg 185

```

```

Db 201 GAAACCGGTCGCGAGTCGTTTCCATTGATGGCGCAACGAATTCGACCCACTCTTCGAA 260
Qy 186 SerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPhe 205
Db 261 GAGCTCTAGACTGTGTGCGAAGCCCGGCGCTCGACACCTGTGATCCACGAAGAC 320
Qy 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225
Db 321 GCCCGCGGCACATGGAATCACTTCGCCACGGGATGCGCTCGACCTCGCTGACCCAG 380
Qy 226 HisIleLeuAlaLysHisGlyValArgGluMetAlaGluInSerGlyMetValAlaThr 245
Db 381 ATCACTGTGTTCAAGCGCACCTCGCGAGGCGCTAAAGCAACACGTCGCCGCCACG 440
Qy 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
Db 441 TTCAATGCCAAGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 500
Qy 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer 285
Db 501 GTCGACATCGCACCGGCAACACGAGTTCGTCGACGCTGACGCGC-----AACAGAGT 554
Qy 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhe 305
Db 555 CGCTGTTCTCTGCACCATCGCGGCTTGCGAATAATACATCCCAAGCTGCTGCCGATG 614
Qy 306 PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys 325
Db 615 TTGCGCCCAACGTAACCTGTTCCGTCGCTTCCTGCGGACACTTCGGCACCGGTCAC 674
Qy 326 CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuGlyAsnSerLysSerGlu 345
Db 675 GTCGAGTGGGGGAGAGAACCGCACCGTTCGCTGCGGCTGCGGCTGCGGCTGCGGCTG 734
Qy 346 GlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrTrpLeuAlaPheSer 365
Db 735 GCATCGCGGTGGAGAACGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 794
Qy 366 AlaIleLeuAlaAlaGly 371
Db 795 GCTAGTCTGCTGTGTGC 812

```

```

RESULT 13
BJ358241/c
LOCUS
DEFINITION
Dictyostelium discoideum cDNA clone dda5622 3', mRNA sequence.
ACCESSION
BJ358241
VERSION
BJ358241.1 GI:19257836
KEYWORDS
Dictyostelium discoideum
SOURCE
Dictyostelium discoideum
ORGANISM
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE
1 (bases 1 to 684)
AUTHORS
Urushihara,H., Tanaka,Y., Kohata,Y. and Shin-i.T.
TITLE
Full length cDNA of Dictyostelium discoideum at the aggregation
stage
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1. .684
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/cloned="dda5622"
/sex="mat A"

```

FEATURES

```

source
1. .684
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/cloned="dda5622"
/sex="mat A"

```

```

/dev stage="Aggregation stage"
/clone lib="Dictyostelium discoideum cDNA library, AF"
BASE COUNT 202 a 103 c 108 g 256 t 9 others
ORIGIN
Alignment Scores:
Pred. No.: 1,45e-25 Length: 684
Score: 316.50 Matches: 76
Percent Similarity: 56.12% Conservative: 34
Best Local Similarity: 38.78% Mismatches: 80
Query Match: 13.32% Indels: 6
DB: 12 Gaps: 4
US-10-098-602A-2 (1-454) x BJ358241 (1-684)
Qy 256 GlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPhe 275
Db 669 GGCTGCTCAGGACACATGCATCAAAATTTCAATTTTAAACACCGTAAGAACNTATTC 610
Qy 276 TyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeu 295
Db 609 CTCGATGAATCCGACCCAAATCATATGCTGCATTTTCAATCATTCGTTGNTGTCGCA 550
Qy 296 LeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArg 315
Db 549 TTATTATTGTTGCTGAGTTTTTACCATTTCGACCAACAATCNATAGTTATAACGT 490
Qy 316 GlnProLeuThrPheAlaProThrLysCysTyrTrpAlaIleAspAsnArgThrSer 335
Db 489 TTATGTCGATGGGTATTGGGCACCAACCAACATGGGAGTGGTAATAGACTGTA 430
Qy 336 AlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGly 355
Db 429 GCATTAAGAAATATCAAAAGGT---GGTAAAGCAACAGATCAGATTTCAGATCAGGA 373
Qy 356 AlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIle 375
Db 372 TCAGATGAAATCCCTTACATTTCAATCGCGCTTCATTTCGCTGCTGCTCTATGCTGC 313
Qy 376 GluGluLysLeuGluLeu---ProProAlaSerGlyAsnValTyrAsn-----App 392
Db 312 ATTAACAAATTAGAATTAACAAACCAACCATCATAGGTAATAGTATGATCTCTATAAA 253
Qy 393 LysGluLeuProGlu---PheProAsnSerLeuGlnAsnAlaThrHisLeuLysGlu 411
Db 252 AAAGACTAGTTGAAAGATTACCAAGATCATTTAGTGTATCACTCACTGATTAATCAAAA 193
Qy 412 SerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAla 431
Db 192 TCAAGATTCGCAAAAGATACCTTGTGGAAGAATTTGTNGATCATTTCTGTTGAACTCGT 133
Qy 432 AsnValGlu-IleAsnGluPheSerLysGlnValThrAspTrpGlu 446
Db 132 CGTTGGGAATAGACACATTCATCAATCAATCAAGTTCATAAATGGGAA 87
RESULT 14
BE636680
LOCUS
DEFINITION
rockefeller.0.366 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to glutamine synthetase (EC
6.3.1.2), mRNA sequence.
ACCESSION
BE636680
VERSION
BE636680.1 GI:9919791
KEYWORDS
EST.
SOURCE
Mastigamoeba balamuthi
ORGANISM
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE
1 (bases 1 to 1138)
AUTHORS
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sansen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

```



```
Qy 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CCATTTTCGCCACCAACATCAATAGTTTAAAGTTTAGTCGATGGTATGGGCACCA 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 324 ThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 ACCACCAACATGGGAATGGATAATAGAATCTAGCATTAGAATTTATCAAGGT--- 177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 GGTAAAGCAACAAGATCAGAAATTCAGAGTCACAGGATCAGATGTAATCCTTACATTCA 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 364 PheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu---Pro 382
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 ATCCCGCTTCATTCGCTCGCTCGTCTCATGCTGTGCTTAACAATTAGAATTAACAACAA 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 383 ProProAlaSerGlyAsnValTyrAsn-----AspLysGluLeuProGlu---PhePro 399
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 AAACCAATCATAGTATAGTTATGATCTCTATAAAAAAGGACTAGTTGAAAGATTACCA 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 400 AsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPhe 419
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 AGATCATTAGCTGAATCAACTGAATTAATCAAAATCAAAAGATTGCAAAAGAAATACCTT 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 420 GlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSer 439
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 GGTGAAGAAATTTGTGATCATTTGTTGAAACTCGTCGTTGGGAATATAGACAATTCAT 477
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 440 LysGlnValThrAspTyrGluLeuAsnGln 449
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 CATCAAGTTTCATAAATGGGAATTAGAAAGA 507
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: December 18, 2003, 01:17:46
Job time : 2974 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 18, 2003, 00:54:51 ; Search time 457 Seconds
(without alignments)
3310.329 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLIRNGKIDTV.....INEFSKQVTDWELNGFNRY 454

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2211978 seqs, 1666101734 residues

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPFO_spool/US10098602/runat_17122003_151240_28154/app_query.fasta_1.647
-DB=Published_Applications_NA -Qfmt=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10098602 @CGN 1 1 353 @runat_17122003_151240_28154
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPELOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description

1	773.5	32.6	1362	15	US-10-156-761-6703	Sequence 6703, Ap
2	773.5	32.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
3	475.5	20.0	5500	15	US-10-062-458-1	Sequence 1, Appli
4	470.5	19.8	1338	10	US-09-738-626-2442	Sequence 2442, Ap
5	470.5	19.8	3309400	10	US-09-738-626-1	Sequence 1, Appli
6	449	18.9	4259	9	US-09-733-383-12	Sequence 12, Appl
7	449	18.9	4259	13	US-10-199-405-12	Sequence 12, Appl
8	432	18.2	1347	9	US-09-815-242-9280	Sequence 9280, Ap
9	427.5	18.0	1359	15	US-10-156-761-5931	Sequence 5931, Ap
10	424.5	17.9	1338	9	US-09-815-242-4678	Sequence 4678, Ap
11	424.5	17.9	1341	9	US-09-815-242-8240	Sequence 8240, Ap
12	424.5	17.9	1341	9	US-09-815-242-8677	Sequence 8677, Ap
13	402	16.9	1356	9	US-09-815-242-6417	Sequence 6417, Ap
14	402	16.9	4176	10	US-09-070-927A-188	Sequence 188, App
15	395.5	16.6	1008	10	US-09-974-300-261	Sequence 261, App
16	328	13.8	722	10	US-09-974-300-4790	Sequence 4790, Ap
17	322	13.6	1431	10	US-09-738-626-3485	Sequence 3485, Ap
18	322	13.6	3309400	10	US-09-738-626-1	Sequence 1, Appli
19	300.5	12.6	918	15	US-10-156-761-200	Sequence 200, App
20	291	12.2	1407	15	US-10-156-761-5982	Sequence 5982, Ap
21	291	12.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
22	287	12.1	1419	9	US-09-815-242-7040	Sequence 7040, Ap
23	287	12.1	1830121	15	US-10-329-960-1	Sequence 1, Appli
24	281.5	11.8	1503	15	US-10-187-267A-56	Sequence 56, Appl
25	281.5	11.8	36321	15	US-10-187-267A-1	Sequence 1, Appli
26	277	11.7	82993	16	US-10-080-170-645	Sequence 645, App
27	265.5	11.2	1410	9	US-09-815-242-7990	Sequence 7990, Ap
28	265	11.2	1446	9	US-09-815-242-7410	Sequence 7410, Ap
29	247	10.4	1410	9	US-09-815-242-9691	Sequence 9691, Ap
30	241	10.1	1410	9	US-09-815-242-6321	Sequence 6321, Ap
31	241	10.1	1410	13	US-10-299-799-2	Sequence 2, Appli
32	201	8.5	372	10	US-09-974-300-4799	Sequence 4799, Ap
33	159	6.7	330	10	US-09-974-300-4800	Sequence 4800, Ap
34	126.5	5.3	607	13	US-10-027-632-275319	Sequence 275319,
35	126.5	5.3	607	14	US-10-027-632-275319	Sequence 275319,
36	114.5	4.8	1419	15	US-10-205-823-151	Sequence 151, App
37	111.5	4.7	48715	13	US-10-238-075-921	Sequence 921, App
38	111.5	4.7	50538	13	US-10-085-959-57	Sequence 57, Appl
39	107	4.5	26668	10	US-09-962-832-222	Sequence 222, App
40	105	4.4	2793	10	US-09-917-800A-1629	Sequence 1629, Ap
41	104.5	4.4	2727	10	US-09-880-107-3748	Sequence 3748, Ap
42	102.5	4.3	858	10	US-09-974-300-1348	Sequence 1348, Ap
43	101	4.3	3567	15	US-10-101-464A-47	Sequence 47, Appl
44	100	4.2	5715	14	US-10-005-691-1	Sequence 1, Appli
45	100	4.2	6057	15	US-10-331-061-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-10-156-761-6703
; Sequence 6703, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6703
; LENGTH: 1362
; TYPE: DNA

QY 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGlu 301

97: CT

DB: _____
TS: _____
Grades: _____

US-10-098-602A-2 (1-454) x US-10-156-761-1 (1-9025608)

Qy 6 AspGluLeuAenAsnLeuIleAArgAsnGlyLysIleAspThrValValLeuAaCysVal 25
Db 8041067 GAGGAGCTGCAGCGCTCGTGGCGGCGGAGATCGACATGTCGTCTCGGTCC 8041126

Qy 26 AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45
Db 8041127 GATATGCAAGGCGCTCCAGGCAAGCGGTTCGCGCGCGCTTCTCTC----- 8041177

Qy 46 GlnLysLysIleSerIleSerThr-----PheValTyAlaValThrIleGlu 61
Db 8041178 GACGAGGTGCTGGAGCAGCGGAGGAGATGCAACTACCTGCTGCGCTGCACACCGAC 8041237

Qy 62 GlyIleAlaGlyGlyTyGluIleSerValAspThrGlyTySerAspCysHis 81
Db 8041238 ATGAACACCGTCGACGGCTATGACATGTCCTCTGGACCGCGGTACGGGACCTCGCC 8041297

Qy 82 LeuCysAlaAspLeuAenSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAla 101
Db 8041298 ATGCATCCGACCTGAGCAGCTGCGCCAGGTGCGTGGAAACGCGGCGACCGCTGCTG 8041357

Qy 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121
Db 8041358 ATCCGCGAC---CTCGCCTGGAACGACGATCGCGGTGGTGGTGGCGCCCGCCGACATC 8041414

Qy 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysGlyLeuPheAlaSerGlu 141
Db 8041415 CTGGTCCGACGCTGACCGGCTGCGCCCTCGGCTACCGCCAGGTGCGGACCGAG 8041474

Qy 142 LeuGluPheAenLeuPheAenGluThrTyLysSerAlaSerGlnLysHisTrpLysAsn 161
Db 8041475 CTGGAGTTTCATCGTCTTCAAGGACACCTACGAGGAGCTGGAGCGCGGATACAAGGG 8041534

Qy 162 LeuLysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerGlyIleGlu 181
Db 8041535 CTCACCGGCGCAACCACTACGACTACTCGTGTCTGGAGCGGACGGATCGAA 8041594

Qy 182 ThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThr 201
Db 8041595 CCCTGTCGCGCGCATCCGCAACGAGATGAGCGCGCGGCTCACCGCTGCGCGCC 8041654

Qy 202 HisProGluPheLeuProSerGlnHisGluLeuAenPheValProAlaAspProLeuThr 221
Db 8041655 AAGGCGAGTCAACCCCGGCGCAGCAGATCGTGTTCGGTAGCAGCGCCCTGCTC 8041714

Qy 222 MetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGlnSerGly 241
Db 8041715 ACCTGCGCAGCAGCGCGCTTACAAGACCGCGCCCAAGGAGATCGCGCCCGGAGGGC 8041774

Qy 242 MetValAlaThrPheMetAlaLysLeuSerThrAlaLeuGlyAenAlaCysHisIle 261
Db 8041775 GTGTGATCATTCTATGGCCCAAGTACCAACGAGCGCGAG---GGCACTCTGCGCACATC 8041831

Qy 262 HisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyArgGlnAsnAspGlu 281
Db 8041832 CACCTGTGCTGGCGGCGCGGAC---GGCACCACCGCATGCGCGCGGCGGACCGCCGC 8041898

Qy 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyValProGlu 301
Db 8041899 ---GGCATGTGCGAGCTCATGCGCAGCTTCTCGCGGCGGAGCTCGCGCGCTGCGGCGAC 8041945

Qy 302 AlaThrTyPhePheAlaSerTyIleAsnSerTyIleAsnArgLeuGlnProLeuThrPhe 321
Db 8041946 TTCTACTTCTGTAGCGGCCCAACATCACTCGTACAGCGGTTCACGCGGCTCTCTTC 8042005

Qy 322 AlaProThrLysCysTyPThrAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn 341
Db 8042006 GCGCGGACCGCGTGGCTGGGCTAGCACAACCGACCTCGCGGCTCGCGGTGCTC--- 8042062

Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyIleAlaAspLeuAsnProTyr 361
Db 8042063 GGCCACCGCGCTCGATCGCTTCGAAACCGGCTCCCGCGGTGAGTCAACCCGAC 8042122

Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 8042123 CTCGCCCTCGCGGACTCGTGGCGCGGCTGTACGCGATCGAGCACAAGCTGGAGCTG 8042182

Qy 382 ProProAlaSerGlyAsnValTyAsnAspLysGluLeuProGluPheProAsnSer 401
Db 8042183 CCCGAGCGCTCGCGGCGCAACGCGCTAC---GCCCGGAGTACGAGCAGCTCCACACACC 8042239

Qy 402 LeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlu 421
Db 8042240 CTGGCGAGCGCGCGCAACTCTGGGAAACAGACCCCAAGGCGCGCTTCGCGCAC 8042299

Qy 422 LysLeuIleLeuHisTyValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGln 441
Db 8042300 GAGTGTGCGCCACTACCGCAACATCGCGCGCTCGACTGAGGCGCTTCGACGCCG 8042359

Qy 442 ValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 8042360 GTGACCGACTGGGAGCTGCGCGCTCTTCGAACGC 8042395

RESULT 3

US-10-062-458-1
; Sequence 1, Application US/10062458
; Publication No. US20030003550A1
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: IZUI, HIROSHI
; APPLICANT: MORIGUCHI, KAYO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMINE BY FERMENTATION AND L-GLUTAMINE
; FILE REFERENCE: 219181USO
; CURRENT APPLICATION NUMBER: US/10/062,458
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: JP 2001-28163
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: JP 2001-162806
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 5500
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (659)..(1996)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2006)..(5200)
; OTHER INFORMATION:
US-10-062-458-1

Alignment Scores:
Pred. No.: 6,94e-50 Length: 5500
Score: 475.50 Matches: 135
Percent Similarity: 47.08% Conservative: 83
Best Local Similarity: 29.16% Mismatches: 174
Query Match: 20.01% Indels: 71
DB: 15 Gaps: 12

US-10-098-602A-2 (1-454) x US-10-062-458-1 (1-5500)

Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34
Db 743 GGACACTTGAAGTCAAGTGGTGGCTCTCTGCGAGAACTAGAGTCTGGCTG----- 793

Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPhe 54
Db 8042395

Db	794	-----GAAGAAGGCATCGGATTTCGAT-----	814
Qy	55	ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp	74
Db	815	-----GGCTCAGCCATTGAGGGCTACGCG-----CGTATCTCGGAAGCGGAC	856
Qy	75	ThrGlyTyrSerAspCysHisLeuCyAlaAspLeuAsnSerLeuHisLeuLeuProTyr	94
Db	857	ACC-----ATTGCCGCCCGCCAGATCCATCGACATTTCCAGGTCTCTCCCACTA	901
Qy	95	SerGluGly---AlaValLeuAlaIleSerAsnProHisAsnPheValThr-----	110
Db	902	GAACGGCGGCATCTCAAAACTCGAGCAGCACGCCCTGTTTTCGGATGTCAGATGCCGGAC	961
Qy	111	SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla	130
Db	962	GGACAGCCATCTTTTCTGACCCCGCCCAAGTGCTGGCAGGACAGGTCCCAACTAGTGCA	1021
Qy	131	AsnLeuIysLeuIysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr	150
Db	1022	GATCAAGGCTTGACCTGCATGATCTCACAGAGATTGAGTCTTATTGTGGTG-----	1072
Qy	151	TyrIysSerAlaSerGlnIlyHisTrpIysAsnLeuIysThr-----	164
Db	1073	-----CAAGCCCTTCGCACCAACGGCATGCCACCTGTG	1105
Qy	165	AlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet	184
Db	1106	CCCACTGACAAAGCGCGGANTATTTCGACCAAGCCACATTCATGAGGCGCGAATTTCCGT	1165
Qy	185	ArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGlu	204
Db	1166	CGAAACGCGATGGTAGCGCTGGAGGAACTCGGCATCCTCTCGAGTTCTCCCAACATGAA	1225
Qy	205	PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp	224
Db	1226	ACTGCACCTGGCGACGACGAAGAAATCGATTACGCCATCGGATGCGTTCACCATGGCGAC	1285
Qy	225	ArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla	244
Db	1286	AACATCATGACCTTCGCGTACATCAAAACAGGTGGCAGGACCAAGGCGTTGGGGCA	1345
Qy	245	ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCyHisIleHisMetSer	264
Db	1346	TCATTTATGCCAAGCCATTCCAAGACATGCAGGCTCCGCCATGCACACGCACATGTC	1405
Qy	265	LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet	284
Db	1406	TTATTTCGGGCGCATACC-----AAGCGGTTCACGATCCAGCATCTTTACATGCTG	1459
Qy	285	SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuIysTyrValProGluAlaThrTyr	304
Db	1460	TCCAAAACCGCAAAACAGTTCATCGCTGGAACTTTGCATCAGCGTCCAGAAATTCACCGCT	1519
Qy	305	PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr	324
Db	1520	GTGACCAACCAAGTGGGTCAATCTCTCAACAGCATCGTGTACGGAACGAGCTCCAACT	1579
Qy	325	LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn-----	341
Db	1580	CGCGCAACCTGGGTGTATCTAATCGTTCTCGCGTGGTTCGTGTTCTCACTACCGTTTG	1639
Qy	342	SerLysSerGluGlyIleAsnValGlnLeuArgIleGlyGlyAlaAspLeuAsnProTyr	361
Db	1640	AATAAGGAGGAGTCCGCCCGGTCGAGGTGGCTCTTCTGATACCGCTTGTAAACCCATAT	1699
Qy	362	LeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGlyLysLeuGluLeu	381
Db	1700	TTGCGTHTTCATGATGCTCGCGCGCTGGTGTGAAGAGTATTAAAGAAAGGTTATGAGCTC	1759
Qy	382	ProProAlaSerGlyAsnVal-----	389
Db	1760	GACGAGCCAGCTGAGGACGATATCTTCCAACTTGAGCTTCGGGAAACGTGCGGCCATGGC	1819

390	TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu	409
1820	TACACAGCAT-----CTGCCAAACAGCCTTGATCAGCAGCATCGCCCAAATG	1864
410	LysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsn	429
1865	GAAGAATCAGAGCTGTGTGCTGACATCCTCGGTGAGCAGCGTTTGTAGTGTCTTCTGCGC	1924
430	AlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln	449
1925	AATAAGTGGCGTGAATCGCGTGTACTACCAAGACGAGATCACTCCGTGGGAGCTCCGAAC	1984
450	GlyPheAsn	452
	:::	
1985	AATCTTGTAT	1993
	:::	

RESULT 4

```

US-09-738-626-2442
; Sequence 2442, Application US/09738626
; Publication No. US20020197605A1
;
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HATASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIORITY APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2442
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2442

```

Alignment Scores:	
Pred. No.:	2.71e-50
Score:	470.50
Percent Similarity:	47.08%
Best Local Similarity:	28.94%
Query Match:	19.80%
DB:	10
	Gaps:
	12
	Indels:
	71
	Mismatches:
	174
	Conservative:
	84
	Matches:
	1338
	Length:
	1338

US-10-098-602A-2 (1-454) x US-09-738-626-2442 (1-1338)

Qy	15	Gly	Lys	Ile	Asp	Thr	Val	Val	Leu	Leu	Ala	Cys	Val	Asp	Met	Gln	Gly	Arg	Leu	Met	Gly	Lys	34
Db	85	GGC	CAC	TTG	AAG	T	CAG	T	GGT	T	TG	T	GG	T	CCT	C	CAG	A	ACT	T	AG	T	CGCTG
Qy	35	Arg	Leu	Thr	Gly	Arg	G	His	Phe	Leu	Gly	Leu	Asp	Gln	Lys	Lys	Ile	Ser	Thr	Phe		54	
Db	136	-----	GA	AG	AG	GC	AT	CG	GA	TC	GC	AT	-----	-----	-----	-----	-----	-----	-----	-----	-----	156	
Qy	55	Val	Tyr	Ala	Val	Thr	Ile	Glu	Gly	Ile	Ala	Gly	Gly	Tyr	Gly	Glu	Ile	Ser	Val	Asp		74	
Db	157	-----	GG	CT	C	AG	CA	TT	AG	GG	GC	TAC	GGC	-----	-----	-----	-----	-----	-----	-----	-----	198	
Qy	75	Thr	Gly	Trp	Ser	Asp	Cys	His	Leu	Cys	Ala	Asp	Leu	A	Asn	Ser	Leu	His	Leu	Leu	Pro	Trp	94


```
Db 199 ACC-----ATTGCCCGCCAGATCCATCGACATTCAGGTCCTCCCACTA 243
Qy 95 SerGluGly---AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110
Db 244 GAACGGGCGATCTCAAACTCGACGACGACCGCTGTTTCGGATGTCACATGCCAGAC 303
Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGluArgLeuAla 130
Db 304 GGACAGCCATCTTTCTCTGACCGCGCAAGTGTGCGCAGGAGTCCAACTAGCTGCA 363
Qy 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150
Db 364 GATGAAGCTTGCATCGATCATCTACACAGAGATTGAGTTCTATTGGTG----- 414
Qy 151 TyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThr----- 164
Db 415 -----CAAGCCTTCGACCAACGAGTCCACCTGTG 447
Qy 165 AlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
Db 448 CCCACTGACACGCGGATATTTCGACCAAGCCACATTCATGAGGCGCGCAATTCGCT 507
Qy 185 ArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
Db 508 CGAAACGCGATGGTAGCGCTGGAGAACTCGGCATCCCTGCGAGTTCTCCACCATGAA 567
Qy 205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224
Db 568 ACTGCATCTGCCAGCAAGAAATCGATTACGCCATGCGGATCGCTCACCATGCCGAC 627
Qy 225 ArgHisIleLeuAlaLysHisGlyValArgLysMetAlaGluGlnSerGlyMetValala 244
Db 628 AACATCATGACCTTCCGCTACATCATGAAACAGGTGGCAAGGACCAAGCGTCGGGCA 687
Qy 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer 264
Db 688 TCATTATGCCCAAGCCATTCCAAAGAACATCGAGCTCGGCCATGCACACGCATGCTCC 747
Qy 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheThrArgGlnAsnAspGluTyrGlyMet 284
Db 748 TTATTTCGGCGGATACC-----AACCGTTTCACCATCGACATCTTACATGCT 801
Qy 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyr 304
Db 802 TCCAAACCGCAAAACAGTTTCATCGTGGATCTTGTCATCAGCTCCAGATTCACCGCT 861
Qy 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324
Db 862 GTGACCAACCAAGTGGTCAATCTCTACAAACGCATCGTGTACGGAAACGAAGCTCCAAC 921
Qy 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341
Db 922 CGCGCAACCTGGGTGATCTAATCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 981
Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyr 361
Db 982 AATAAGGAGGAGTCCGCGCGGTGGAGTGGCTCTCTGATACCGCTTGTAAACCATAT 1041
Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 1042 TTGCGGTTTTTCAGTGATGCTCGCGCTGTTTGAAGAGCATTAAGAGAGGTTATGAGTCT 1101
Qy 382 ProProAlaSerGlyAsnVal----- 389
Db 1102 GACGAGCGCTGAGGAGATATCTCAACTTTCGCGGAACTCGCGCATCGGCTGGC 1161
Qy 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409
Db 1162 TACAACGAT-----CTGCCAGAGAGCTTGTATCAGGACACTCGCGCAATG 1206
Qy 410 LysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuLeuHisTyrValala 429
Db -----
```

```
Db 1207 GAAAGTCAGAGCTTGTGTGATCATCTCGGTGAGCAGCAGCTTTTGTAGTTTTTCTGCGC 1266
Qy 430 AlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
Db 1267 AATAAGTGGCGTGAATGGCGTGACTACCAAGACGACATCACTCCGTGGAGCTCCGAAC 1326
Qy 450 GlyPheAsn 452
Db 1327 AATCTTGAT 1335
```

RESULT 5

```
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
```

```
Alignment Scores:
Pred. No.: 1,94e-44 Length: 3309400
Score: 470.50 Matches: 134
Percent Similarity: 47.08% Conservative: 84
Best Local Similarity: 28.94% Mismatches: 174
Query Match: 19.80% Indels: 71
DB: 10 Gaps: 12
```

US-10-098-602A-2 (1-454) x US-09-738-626-1 (1-3309400)

```
Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34
Db 2364071 GGGCACTTGAAGTCAGTGGTGTGGCTCTCGCAGAACTAGAGTCGCTGTTG----- 2364021
Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPhe 54
Db 2364020 -----GAAGAAGGCATCGGATTCGAT----- 2364000
Qy 55 ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp 74
Db 2363999 -----GGCTCAGCCATTTAGGGCTACGG-----CGTATCTCGAAGCCGAC 2363958
Qy 75 ThrGlyTyrSerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrp 94
Db 2363957 ACC-----ATTGCCCGCCAGATCCATCGACATTCAGGTCCTCCCACTA 2363913
Qy 95 SerGluGly---AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110
Db 2363912 GAAGCGGGCATCTCAAAACTGCAGGCAGCAGCGCTGTTTTCGATGTCACGATGCCAGAC 2363853
```

```
QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
Db 2363852 GGACACCCATCTTTCTGACCCGCGCAAGTCTCGCGAGGAGGTCCAACTAGCTGA 2363793
QY 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150
Db 2363792 GATGAAGGCTTACCTGCGATGATCTCACCAGAGATTGAGTTCTATTGTGTTG----- 2363742
QY 151 TyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThr----- 164
Db 2363741 -----CAAAGCCTTCGCACCAACGAGTGCACCTGGT 2363709
QY 165 AlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
Db 2363708 CCCACTGACACGCGCGATATTTCGACCAACGACATTCATGAGCGCGGATTTCCGT 2363649
QY 185 ArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
Db 2363648 CGAAACGCGATGATGCGTGGAGGAACCTCGGCATCCCTGTCGAGTTCTCCACCATGAA 2363589
QY 205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224
Db 2363588 ACTGCACCTGGCGAGCAAGAAATTCGATTTACGCCATCGCGATCGCTCACCATGGCCGAC 2363529
QY 225 ArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla 244
Db 2363528 AACATCATGACCTTCGCTACATCATGAACAGGTGCGAAGGACCAAGCGTCGGGCA 2363469
QY 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer 264
Db 2363468 TCATTTATGCCCAAGCCATTCCAAAGAACATGCGAGGTCCGCCATGCACACGACATGTCC 2363409
QY 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheThrAspGlnAsnAspGluThrGlyMet 284
Db 2363408 TTATTTGGGGCGGCTACAC-----AACGCGTTCCAGATCCAGACCATCTTACATGCTG 2363355
QY 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysIleValProGluAlaThrTyr 304
Db 2363354 TCCAAACCGCAAAACAGTTCATCGCTGGAATCTTCATCATCAGCTCCAGAAATTCACCGCT 2363295
QY 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324
Db 2363294 GTGACCAACACGATGGGTGAATTCCTACAAACGATCGTGTACGAAACGAAAGCTCCAACT 2363235
QY 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341
Db 2363234 CGCGCAACCTGGGGTGATCTAATCGTCTCGCCGTGCTGCTGCTTCTACCTACCGTTTG 2363175
QY 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyr 361
Db 2363174 AATAAGGAGGAGTCGCGCGCGGTGGAGGTGCGTCTTCCTGATACCGCTTGTAAACCATAT 2363115
QY 362 LeuAlaPheSerAlaIleIleAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 2363114 TTGGCGTTTTCAGTGATGCTCGCGGTGTTTGAAGAGCATTAAGAGAGTTATGAGCTC 2363055
QY 382 ProProAlaSerGlyAsnVal----- 389
Db 2363054 GACGACCGAGCTGAGGACGATATCTCCAACTTGAGCTTCGCGGAACGTCGCGCCATGGGC 2362995
QY 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409
Db 2362994 TACAACGAT-----CTGCCAAGCAGCTTCATCAGGCACTGCGCCCAAATG 2362950
QY 410 LysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsn 429
Db 2362949 GAAAGTCAGAGCTTGTGTGTCATCTCCTCGGTGAGCAGCTTTTTCAGTTTTTCTTTCGCG 2362890
QY 430 AlaAlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
Db 2362889 AATAAGTGGCGTGAATGGGTGACTACCAAGACGATCATCTCCGTGGGAGCTCCGAAC 2362830
QY 450 GlyPheAsn 452
```

Db 2362829 AATCTTGAT 2362821

RESULT 6

```
US-09-733-383-12
; Sequence 12, Application US/09733383
; Patent No. US20010014467A1
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjan
; APPLICANT: Rhodes, Michael J. C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/733,383
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 09/155,183
; PRIOR FILING DATE: 1997-03-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 4259
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-09-733-383-12
```

```
Alignment Scores:
Pred. No.: 1,258-46 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 9 Gaps: 6
```

US-10-098-602A-2 (1-454) x US-09-733-383-12 (1-4259)

```
QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db 417 ATGAGTAACAACCTCGACCACTACCGATTGGTTGAAAGACCACAAGATCATCAGAAGTC 476
QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 477 GAATGCATGATTGGCGACTTGACCGGATCACCCCGCGCAAGATCTCGCAACCAACAAG 536
QY 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 537 TTCAITTCGCAAAAAGGATGCGCTCCCGAGAGTGTCTGTTGCAGACAGTACGCGGC 596
QY 61 GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
Db 597 GACTATGTCGAAGACGACATCTAT--TACGAACCTGCTCGACCCGCGCAGATCGACATG 653
QY 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluAlaValLeu 100
Db 654 ATCTCGCGCCCGCAGCAGAACGCGGTGTTCTCTGTCATGGCCATCGAGCGCCGCG 713
QY 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 714 CAGGTGATTTCAGACACCTACGACAGCGGCGCAACCCGATCGAGTGTGCGCACCAAC 773
QY 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140
Db 774 GTCCCTCAAGAAAGTCTCAAACTCTATTCCGACAAAGGGCTGGCAGCGATGCGCGCG 833
QY 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 834 GAAATGGAGTTTCTACCTG-----ACCAAGCGCAGTGCACGCCGATTC----- 878
QY 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
Db 879 -----CCATTGCAACCCCGGTGGCCGCTTCGGACGTCGCGAATTCGTCGCCAA 929
QY 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
```

Db 930 TCGTTCTATCGAAGCGCCCAAGAAATTCACCGCGTGTTCGAGACGCTACGACTGG 989
Qy 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 990 TCGCACTGCAGGAGCTGATCTCGATACGCTGATCCACGAGCGCGCGCGCATG 1049
Qy 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230
Db 1050 GAAATCACTTCGCTCAGCGCGCGCGCTGTCCCTGGCGGACCATCTCGTGTTCAG 1109
Qy 231 HisGlyValArgGluMetAlaGluLysSerGlyMetValAlaThrPheMetAlaLys 250
Db 1110 CGCACCATGCGCGAGCGCGCGCTCAAGCACAACGTCGCGCGCATCTATCGCAAGCG 1169
Qy 251 SerSerThrAlaLeuGluAlaGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
Db 1170 ATGACCGCGCGCTGCGAGCGCATCATCTGACACGATCATCATCATCATCATCAT 1229
Qy 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
Db 1230 GCGAAGAACGCTTTC-----TCCAATGAAGACGGGAGCATGAGCGAGTTGTTCTCAAC 1283
Qy 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrPhePheAlaSerTyrIle 310
Db 1284 CACATCGCGCGCTGCGAAGAAATTCATCCCTGAACGTGCGCGCTGTTCGCGCCCAACGTC 1343
Qy 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrAlaIle 330
Db 1344 AACTCGTCCGCGCTCTCTCGCGACATTCGCGCGCGGTGAACGTGAGTGGGGGAA 1403
Qy 331 AsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlyIleAsnValGlu 350
Db 1404 GAAACCGCTACCGTGGCGCTGCGGTGCGGATCGCGCCCTCAAAACCGTGGGTGAA 1463
Qy 351 LeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAla 370
Db 1464 AACCGCTGCGCGGTGCGGACCAACCGCTGCGGATGCGGAGTGGCGAGCTGCTGTC 1523
Qy 371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390
Db 1524 GGCTACATCGCATGTCGAGAGTATCAACCAAGCGCGCTGCTGGGTGCTGGTGTAC 1583
Qy 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLys 410
Db 1584 GAGCGCGCAACCTG---CGTCTCGCTGACCATCGAAGACGCTCTGGAACGATGAA 1640
Qy 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAla 430
Db 1641 AACAGCAAGACCATCGAATAATCTGGGTCAACAATTCATCATCTGCTGCTGCTGCT 1700
Qy 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTyrGlu 446
Db 1701 AAGCGCGCGAGCATGAATACTTCAAGCGGTGATCATCTGAGTATGGAA 1748

RESULT 7

US-10-199-405-12

Sequence 12, Application US/10199405

Publication No. US2003016751A1

GENERAL INFORMATION:

APPLICANT: Narbad, Arjan

APPLICANT: Rhodes, Michael J.C.

APPLICANT: Gasson, Michael John

APPLICANT: Walton, Nicholas John

TITLE OF INVENTION: PRODUCTION OF p-HYDROXYBENZOIC ACID

FILE REFERENCE: 20747/103

CURRENT APPLICATION NUMBER: US/10199,405

CURRENT FILING DATE: 2002-07-17

PRIOR APPLICATION NUMBER: GB96/06187

PRIOR FILING DATE: 1996-03-23

PRIOR APPLICATION NUMBER: PCT/GB97/00809

PRIOR FILING DATE: 1997-03-24

PRIOR APPLICATION NUMBER: 09/155,183

PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 09/733,383
PRIOR FILING DATE: 2000-12-07
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 4259
TYPE: DNA
ORGANISM: Pseudomonas fluorescens
US-10-199-405-12

Alignment Scores:

Pred. No.: 1,25e-46 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 13 Gaps: 6

US-10-098-602A-2 (1-454) x US-10-199-405-12 (1-4259)

Qy 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db 417 ATGATTAACAACTCGACAGCTCACCGATTGGTTGAAGACCAAGATCAGAGATC 476
Qy 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 477 GAATGTCATGATTGGCGACTTACCGGGATCACCGCGGCAAGATCTCGCAACCAACAAG 536
Qy 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 537 TTCATTGCCGAAAGGCGCATCGCGCTGCCGAGAGTGTCTGTTCAGACAGATCGCGGC 596
Qy 61 GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
Db 597 GACTATGTCGAAGACGACATCTAT---TAGAACTGCTCGACCGCGCGCATCGACATG 653
Qy 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrPheSerGluGlyAlaValLeu 100
Db 654 ATCTGCGCGCGCGCGACGAGCGGTGTTCTCTGTCGATGGGCCATCGAGCGCGCG 713
Qy 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 714 CAGGTGATTCACGACACCTACGACGAGGCAACCGCATCGAGCTGTGCGCACGCAAC 773
Qy 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLysGlyLeuPheAlaSer 140
Db 774 GTCTTCAAGAAAGTCTCAAACTCTATTCGCAAGGGCTGGCAGCGCGATCGTGGCGCG 833
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 834 GAAATGGAGTTCTACCTG-----ACCAAGCGCAGTGCAGCGCGGATTAC----- 878
Qy 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
Db 879 -----CCATTGCAACCGCGGTGGCGCTTCGCGAGCTCGGAAATCGGTGCGCAA 929
Qy 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
Db 930 TCGTTCTCTATCGAAGCGCGCAAGATTCGACCCGCTGTTTCGAAGACGCTCTACGACTGG 989
Qy 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 990 TCGCACTGCAGGAGCTGGATCTCGATACGCTATCCACGAGACGCGCGCGCATG 1049
Qy 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230
Db 1050 GAAATCACTTCGCTCAGCGCGCGCGCTGTCCCTGGCGGACCATCTCTGTTGTTCAAG 1109
Qy 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
Db 1110 CGCACCATGCGCGAGCGCGCTCAAGCAACAGTGGCGCGCGCATCTGTTGCGCAAGCG 1169

```
QY 251 SerSerThrAlaLeuGlyAenAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
DB 1170 ATGACGGCGGAGCTGGCAGCGCATGCACCTGCACAGAGCATCATCGATCGAGACC 1229
QY 271 GluLeuAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
DB 1230 GGCACGAACGCTCTTC-----TCCAATGAAGCGGAGCATGAGCCAGTTGTTCTCTCAAC 1283
QY 291 TrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
DB 1284 CACATCGCGCGCTCGAGAAATTCATCCCTGAACTGCTGCCGCTGTTCGCGCCCAACGTC 1343
QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIle 330
DB 1344 AACTCGTTCGCGCGCTTCCTGCGCGGACATTCGCGCGCGGTGNACTCGAGTGGGCGGAA 1403
QY 331 AspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu 350
DB 1404 GAAACCGTACCTGGCGCTGGCGGTGCGCGATGCGCGGCCCTCAAAACCGTGGGTGGAA 1463
QY 351 LeuArgIleGlyAlaAspLeuAsnProTyrIleAlaPheSerAlaIleIleAlaAla 370
DB 1464 AACCGCTGCGCGGTGCGCGCAACCGCTACCTGCGGATTCGCGCGCGCTGCTGTGTC 1523
QY 371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390
DB 1524 GGTACATCGCGATGCTGAGAGTATCAACCCGCGCGCTGCTGTGGTGGTGGTGTAC 1583
QY 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLys 410
DB 1584 GAGCGCGCAACCTG---CGTCTGCGCTGACCATCGAAGCGCTCTGGAAGCATGGAA 1640
QY 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAla 430
DB 1641 AACAGCAAGACCATCGAGAAATACCTGGGTCACAACTTCATCACTGGCTGCGCGGTC 1700
QY 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446
DB 1701 AAGCGGCGGAGCATGAAACTTCAAGCGCGTATCAGCTCATGGGAA 1748
```

RESULT 8

```
US-09-815-242-9280
; Sequence 9280, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9280
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1347)
US-09-815-242-9280
```

```
Alignment Scores:
Pred. No.: 2,79e-45 Length: 1347
Score: 432.00 Matches: 128
Percent Similarity: 44.80% Conservative: 83
Best Local Similarity: 27.18% Mismatches: 220
Query Match: 18.18% Indels: 40
DB: 9 Gaps: 10
```

US-10-098-602A-2 (1-454) x US-09-815-242-9280 (1-1347)

```
QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
DB 1 ATGCCAATCACAGCTGCAGATATTCGTCGTAAGTCAAGGAAAGAAATGTTACCTTTAT 60
QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
DB 61 CGTCTTATGTTCTCAGATATTTTGGGAACCATGAAACGTCGAAATTCCT----- 111
QY 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla 57
DB 112 -----GCTACAGATGAACAGTTAGATAAGGCTTGTGCAACAGAGTTATGTTGATGA 165
QY 58 ValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyr 77
DB 166 TCTTCTATGAAGTTT-----GTACGTATCAATGAG 198
QY 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGlu--- 96
DB 199 TCGGATATGTTACTTGTACCGGACTTGGATACATGACAGTCTTCCCTTGGGAGATGAA 258
QY 97 ---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115
DB 259 AATCGAAGTGTTCAGGCTGTATCTGTATGCTATACAACAGAGGTGAACCATTTGCG 318
QY 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLys 135
DB 319 GGTGACCTCGGTGTAATTTGAAACGAGCTCTTCGTACATGGAAGAGTTCGATTCAAA 378
QY 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThrTyrLys 152
DB 379 TCCTTCAACCTTGGTCCAGAGCCAGATTCCTCTATTTAAGTTGGATGAAATGGGAC 438
QY 153 SerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMet 172
DB 439 CCAACACTTGAAGTGAATGACAAGGT-----GGCTACTTT 474
QY 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192
DB 475 GATTTGGCACCTTACTGACCTTCGCGCAACACACACGTCGTGAGATTGTGTAATTCCTTGACC 534
QY 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
DB 535 AAAATGGGATTTGAAGTAGAAGCGGATTCACCGAGGTTGCCGTTGGACGATGAGATT 594
QY 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
DB 595 GACTTTAAGTACGATGAAGTTCTCGGTGCTTGTGATAAGATTCAAAATCTTTAAGCTTGT 654
QY 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSer 252
DB 655 GTTAAACCATTTGCTCGCAACACACGAGCTTTACGCAACATTTATGCGGAGGCGCAAAATTT 714
QY 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLys 272
```

```

Db      715 GGTATGCTGGATCAGGTATGCACTGTAATATGCTCTTTGTTGATGCAGAA---GGAAAT 771
Qy      273 AsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet-----SerThrLeuAlaArgAsn 290
Db      772 AACGCTCTTGTATCCAAATGATCAAAAGATGCGAGTGTGTCAGAAACAGCTTACCAT 831
Qy      291 TrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
Db      832 TTCCTAGGCGGTTTGATCAAGCATGCTTACAACTATATGCGCATCATCAACCCACACGTT 891
Qy      311 AsnSerTyrLysArgGlnProLeuThrPheAlaProThrLysCysCysTyrPheAlaIle 330
Db      892 AACCATCAACAAACGTTTCCAGGTTATGAGCGCTGTTTACATTGCTTGGGCTGGT 951
Qy      331 AspAsnArgThrSerAlaPheArgLeuLysAsnSerLysSerGluGlyIleAsnValGlu 350
Db      952 CGTAACCGCTTGCACACTTGTGGCGTACTGCTTACGCTGATGGTATGGAACTCGTCTTGAG 1011
Qy      351 LeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAla 370
Db      1012 TTGGCTTCAGTGGATCCCAATGGCAACCTTACGTTGCTATGGCTGTTCTTTTGGAACTT 1071
Qy      371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390
Db      1072 GGTTTGTATGTTATGAAATAAATAACGAGCACCAGCTCTCTATCGAAGAAATATCTAC 1131
Qy      391 -----AsnAspLysGluLeuProGluPheProAsnSerLeuGln 403
Db      1132 ATCATGACAGCAGAGAGCGCAAGAGCTGGTATTACAGACCTTCCATCACTCTTCAC 1191
Qy      404 AsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeu 423
Db      1192 AACGCTTTGAAAGCTTTGACAGAGATGAAGTGGTTAAAGCTGCTCTCGGAGATCACATC 1251
Qy      424 IleLeuHisTyrValAsnAlaAsnValGluIleAsnGluPheSerLysGlnValThr 443
Db      1252 TACACTAGCTTCTTGAAGCCAAACGAAATCGAATGGGCAAGTTATGCAACCTTCGTTTCA 1311
Qy      444 AspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db      1312 CAATGGGAATTTGATAATTATTATAGACCTTTAC 1344

```

RESULT 9

```

US-10-156-761-5931
; Sequence 5931, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5931
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1359)
US-10-156-761-5931

```

Alignment Scores:

```

Pred. No.: 1,09e-44 Length: 1359
Score: 427.50 Matches: 127
Percent Similarity: 47.52% Conservative: 74
Best Local Similarity: 30.02% Mismatches: 167
Query Match: 17.99% Indels: 55
DB: 15 Gaps: 12

US-10-098-602A-2 (1-454) x US-10-156-761-5931 (1-1359)

Qy      61 GluCluIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyr---SerAsp 79
Db      139 GAGGGCATCGCTTCGACGGCTCGGCATCGAGGGCTTCGCCCGCTGATACGAGTCCGAC 198
Qy      80 CysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAla--- 98
Db      199 ATGATCCCAAGCGGACCGCTCGACGTTCCAGGTCTCTCGCTGGCGCGCGAGGCCCCC 258
Qy      99 -----ValLeuAlaIleSerAsnProHisAsnPheValThrSerGluPro 113
Db      259 GGACACGCCCGCATGTTCTGCGACATCTCATGCGG-----GACGGCTCCCGC 306
Qy      114 LeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLys 133
Db      307 TCCTTCGCGGACCGCGCTAGCTCTCAAGCGCGCTCGCCAAAGACCTCCGACCTGGGC 366
Qy      134 LeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSer 153
Db      367 TTCACCTTACACCCACCGGAGATCGAGTCTTCTCTGCTGAAGGAC----- 414
Qy      154 AlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnPro----- 167
Db      415 -----AAGCGCTGGACGGCTCGCGCCGCGCCCGCGGCGGACAACTCC 456
Qy      168 -----HisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
Db      457 GGCTACTTCGACCAACCGCGCAGACGTC-----GGCATGGAC---TTCCGC 501
Qy      185 ArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
Db      502 CGCCAGCGGATCACCATGCTGGAGTCGATGGGATCTCGGTGCGAGTCTTCCACACGAG 561
Qy      205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224
Db      562 GCGCGCGCGCGCGCAGCAGAGATCGACCTCGGCTACGCCCGCGCCCTCTCGACCGCGGAC 621
Qy      225 ArgHisIleIleAlaLysHisGlyValArgLeuMetAlaGluGlnSerGlyMetValAla 244
Db      622 AACATCATGACGTTCCGCTGTCATGAAGCAGGTGGCGCTGGAGCAGGCGGTGCAGGCG 681
Qy      245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer 264
Db      682 ACCTTCATCGGAGCGGTTCTCGGAGTACCGGGGAGCGGCGATGACACCCACCTCTCG 741
Qy      265 LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet 284
Db      742 CTC-----TTCGAGGCGGACCGCAACGCGTTCTACGAGTCCGGCTCGGAGTACCAC 795
Qy      285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyr 304
Db      796 TCCAAGGTTCGCGCGCTCTTTCATCGCGGCTGCTGAAGCAGCGCGGAGATCTCGGCC 855
Qy      305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr----- 320
Db      856 GTCACCAACAGTGGGTCAACTCTTCAAGCGCATCTTGGGGCGGCTCGGAGCGGCGGCA 915
Qy      321 -----PheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSerAla 336
Db      916 GCGCGCGCGCGCGGAGGCTCGCTGATCTGCTGGGGCCACAAACCGCTCCGCACTG 975
Qy      337 PheArgLeu-----CysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle 353
Db      976 GTCGCGCTCCCGATGTACAAAGCCGCGGCGGCTCGGCGCGGTGCGAGGTCGCTCC 1035

```

QY 354 GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSer 373
 Db 1036 CTGAGCTCGGGCGGACCGGTACCTGGCGTACCGCTGCTCTCGCGGGCGCTCAAG 1095
 QY 374 GlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr-----Asn 391
 Db 1096 GGATCGAGGAGGCTACGAGCTCCCGCGGGCGCGACGACGCTGTGGCGCTCTCG 1155
 QY 392 AppLysGlu-----LeuProGluPheProAsnSerLeuGlnAenAlaThr 406
 Db 1156 GACGCGGAGCGCGCGATGGCATCGAGCCCTCGCGGAGAACCTCGCGGAGGCGCTG 1215
 QY 407 HisLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHis 426
 Db 1216 GCATCTAGCGCGGAGCGAGCTGGCGCGGAGCCCTCGCGGAGGAGCGCTTCGACTTC 1275
 QY 427 TyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446
 Db 1276 TTCCTCGCAACAGAGCAGGAGTGGGAGGAGTACCGCTCGGAGGTGACGGCTTCGAG 1335
 QY 447 LeuAsnGln 449
 Db 1336 CTGCGGAAG 1344

RESULT 10

US-09-815-242-4678
 ; Sequence 4678, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4678
 ; LENGTH: 1338
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-4678

Alignment Scores:
 Pred. No.: 2,61e-44 Length: 1338
 Score: 424.50 Matches: 123
 Percent Similarity: 45.84% Conservative: 92
 Best Local Similarity: 26.23% Mismatches: 211
 Query Match: 17.87% Indels: 43
 DB: 9 Gaps: 9

US-10-098-602A-2 (1-454) x US-09-815-242-4678 (1-1338)

QY 2 ThrIleThrTyrAspGluLeuAsnAsnLeuIleAeGAsnGlyLysIleAspThrValVal 21
 Db 13 ACTTTTCATTAAGACGACATTCGTAAATTTGCAGAGAGGAAATGTAAGATATTAAAG 72
 QY 22 LeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPhe 41
 Db 73 TTCAATTCACCTGATATTTTAGGAACAATTAATAATGTTGAAGTGCCTGTAAGCCAATTA 132
 QY 42 LeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValThrAlaValThrIleGlu 61
 Db 133 -----GAAAAAGTACTTGCATAACGAAATGATGTTTTCACGGTTCCTCTATCGAA 180
 QY 62 GlyIleAlaGlyGlyTyrGluLysSerSerValAspThrGlyTyrSerAspCysHis 81
 Db 181 GGTTC-----GTACGTATCGAAGATTCAGATATGTATC 213
 QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrPheSer-----GluGlyAlaVal 99
 Db 214 TTACATCCAGATTAGATACCTTGGGTAATCTTCCCATGGACTGTGGACAAGTAAAGTT 273
 QY 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArg 119
 Db 274 GCACGTTTAAATTTGTGATGTATATAAAACAGATGTTACACCATTTTCAAGGGGATCCTCGT 333
 QY 120 ValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeu---Phe 138
 Db 334 GCAAACTTAAACGCTGTTTAAAGAAATTTAGGCTTACAGACTTTAAACCTTA 393
 QY 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158
 Db 394 GGGCCTGGAACAGAGATCTTCTGTTTAAAGTTGGATGAAAA----- 435
 QY 159 TrpLysAsnLeuLysThrAlaGlnProHisHisGln-----TrpMet 172
 Db 436 -----GGGAACCAACTTTAGAACTTAAATGATGATGATGATGATATTC 477
 QY 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192
 Db 478 GACTTAGCACCCTACGATTTAGGTGAAAACCTCGCTCGTGATATTGTTTAAAGATTAGAG 537
 QY 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
 Db 538 GATATGGGCTTCGATATTGAAGCTAGTACCATGAAGTTGCCCTGGTCAACATGAAT 597
 QY 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
 Db 598 GACTTTAAATATGCAGATGCTGTTACAGCATGTGATAACATCAACATTTAAATGGTT 657
 QY 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252
 Db 658 GTTAAACCAATCGCACGTAACATTAATTTACACGCAACATTTATGCTTAACCATTTAT 717
 QY 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGlyLys 272
 Db 718 GGTGTGAATGGTAGCGGTATGCACTTTAAAGTTTCATTA-----TTCAAGAGTAAAGAA 771
 QY 273 AsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIle 292
 Db 772 AATGCATTTCTTCGATCCAAATACTGAAATGGGCTTAAACGGAACATGCATATCAATTTACA 831
 QY 293 AlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSer 312
 Db 832 GCAGGTGTACTTAAATAATGCACGCGGATTTACTGCTGTATGTAAACCCGTTAGTAAACTCA 891
 QY 313 TyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsn 332
 Db 892 TATAAACGTTTAGTACCTGGTTATGAAGCACCATGTTATATGTCATGGAGTGGTAAAC 951
 QY 333 ArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArg 352
 Db 952 CGTTCCACCAATTAATCCGTGTACCATCTTCAAGAGGATTATCAACTCGTATCGAAGTACGT 1011

```

Qy 353 IleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIle 372
Db 1012 TCAGTAGATCCAGCTGCAAAACCCATACATGCGGTAGCTCAATCTTAGAAGCTGGACTA 1071
Qy 373 SerGlyIleGluGlyLeuLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db 1072 GATGGTATTAATAATAAATTAAGATTCAGAACCCAGTTAACCCAAAATATTACGAAATG 1131
Qy 391 AsnAspLysGlu-----LeuProGluPheProAsnSerLeuGlnAsnAla 405
Db 1132 AACCGTGAAGAACGTGAAGCAGTAGGATTCAGACCTTACCTTCAACACTTATACGTGCA 1191
Qy 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeu 425
Db 1192 TTAAGAGCAATGCGGTGAATGCAAGTATTATAAAGCTTTAGGAAATCATATCATATAAT 1251
Qy 426 HisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTyr 445
Db 1252 CAATTTATTAATTCAAATCAATTAATGAATGGGATTACTACAGAACTCAAGTATCTGAATGG 1311
Qy 446 GluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1312 GAAGAGATCAGTACATGAACCAATAT 1338

```

RESULT 11

```

US-09-815-242-8240
; Sequence 8240, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8240
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1341)
US-09-815-242-8240

```

```

Alignment Scores:
Pred No.: 2,62e-44 Length: 1341
Score: 424.50 Matches: 123
Percent Similarity: 45.84% Conservative: 92
Best Local Similarity: 26.23% Mismatches: 211

```

```

Query Match: 17.87% Indels: 43
DB: 9 Gaps: 9
US-10-098-602A-2 (1-454) x US-09-815-242-8240 (1-1341)

Qy 2 ThrIleThrTyrAspGluLeuAsnAsnLeuIleAsgAsnGlyIleAspThrValVal 21
Db 13 ACTTTCATAAGACGACATTCGTAATTTTCGACAGAGGAGAAATGTAAGATATTAAAGA 72
Qy 22 LeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPhe 41
Db 73 TTCAATTCCTGATATTTTAGGAACCAATTAATAATGTTGAAGTGCTGTGAAGCAATTA 132
Qy 42 LeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGlu 61
Db 133 -----GAAAAAGTACTTGTATAACGAAATGATGTTTGACGGTCTCTCTATCGAA 180
Qy 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
Db 181 GGTTC-----GTACGTATCGAAGAAATCAGATATGTCAT 213
Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuProTyrSer-----GluGlyAlaVal 99
Db 214 TTATCCAGATTTAGATACCTTGGGTAACTTCCATGGACTGCTGGACAAAGGTAAGATT 273
Qy 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArg 119
Db 274 GCAGGTTTAATTTGTGATGTATATAAACAGATGTTACACCATTTGAAGGGGATCCTCGT 333
Qy 120 ValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLysGlyLeu---Phe 138
Db 334 GCAAACTTAAACGTTGATTATAAAGAAATGGAAGATTTAGGCTTCACAGACTTTAACTA 393
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158
Db 394 GGGCCTGAACCAAGAAATTCCTTGTGTTAAGTTGGATGAAAAA----- 435
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisGln-----TrpMet 172
Db 436 -----GGGGAACCACTTTAGAACTTAATGATGATGATGATTTTC 477
Qy 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192
Db 478 GACTTAGACCTACAGATTTAGGTGAAACTGCTGCTGATATTTGTTAGAAATTAGAG 537
Qy 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
Db 538 GATATGGCTTCGATATTTGAAGCTAGTACCACCAATGTCCTCCCTGTCACATGAAT 597
Qy 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
Db 598 GACTTTAAATATCGAGATGCTGTACAGCATGTGATAACATCCAAACATTTAAATGCTT 657
Qy 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252
Db 658 GTTAAACAAATCGCACGTAACATAATTTACACGCAACATTTATGCTAAACCATTTATTT 717
Qy 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLys 272
Db 718 GGTGTGAATGGTAGCGGTATGCACTTTAAAGTTTCATTA-----TTCAAGAGGTAAGAA 771
Qy 273 AsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIle 292
Db 772 AATGCATTTCTCGATCCAAATACTGAATGGGCTTAACGGAAACTGCATATCAATTTACA 831
Qy 293 AlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSer 312
Db 832 GCAGGTGTACTTAAAAATGCACCGCGATTTACTGCTGTATGTAACCGCTAGTAAACTCA 891
Qy 313 TyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsn 332
Db 892 TATAAACGTTTAGTACCTGTTATGAAGCACCACCATGTTATATTTCATGCGAGTGGTAAAAAC 951

```

QY 333 ArgThrSerAlaPheArgLeuGluCysAenSerLysSerGluGlyIleAenValGluLeuArg 352
 Db 952 CGTTCCACCTTATCGTGTACCATCTTCAAGAGGATTATCAACTCGTATCGAGTACGT 1011
 QY 353 IleGlyGlyAlaAspLeuAenProTyrLeuAlaPheSerAlaIleIleAlaGlyIle 372
 Db 1012 TCAGTAGATCCAGCTGCAAAACCCATACATGCGCTTAGCTGCAATCTTAGAAGCTGACTA 1071
 QY 373 SerGlyIleGluGlyLeuGluLeuProProAlaSerClyAenValTyr----- 390
 Db 1072 GATGGTATTAAAAATTAATTAAGTTCAGAACCACTTAACCAAAATATTTACGAAATG 1131
 QY 391 AsnAspLysGlu-----LeuProGluPheProAenSerLeuGlnAenAla 405
 Db 1132 AACCGTAGAAGCTGAGCAGTAGGCATTCAGACTTACCTTCAACACTTTATATCTGCA 1191
 QY 406 ThrHisLeuLeuGlySerLysMetLeuAenLysThrPheGlyGlyLysLeuLeu 425
 Db 1192 TTAAGGCAATGCGTGAAATGAAGTATTAAAAAGCTTTAGGAAATCATATCTATAAT 1251
 QY 426 HisTyrValAsnAlaAenValGluIleAenGluPheSerLysGlnValThrAspTyr 445
 Db 1252 CAATTTATTATTAATCAAAATCAATTTGAATGGATTACTACAGAACTCAAGTATCTGAATG 1311
 QY 446 GluLeuAenGlnGlyPheAenArgTyr 454
 Db 1312 GAAAGAGATCAGTACATGAGCAATAT 1338

RESULT 12

US-09-815-242-8677
 ; Sequence 8677, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8677
 ; LENGTH: 1341
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus aureus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1341)
 US-09-815-242-8677

Alignment Scores:

Pred. No.: 2.62e-44 Length: 1341

Score: 424.50 Matches: 123
 Percent Similarity: 45.84% Conservative: 92
 Best Local Similarity: 26.23% Mismatches: 211
 Query Match: 17.87% Indels: 43
 DB: 9 Gaps: 9
 US-10-098-602A-2 (1-454) x US-09-815-242-8677 (1-1341)
 QY 2 ThrIleThrTyrAspGluLeuAenAenLeuIleAenGlyLysIleAspThrValVal 21
 Db 13 ACTTTCATCAAGAGCATTCGTAAATTTGCAGAGAGGAAATATTAAGATATTAGA 72
 QY 22 LeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPhe 41
 Db 73 TTACAATTCACGTATATTAGGAACAATTAATAATGTTGAAGTCCTGTAGACCAATTA 132
 QY 42 LeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGlu 61
 Db 133 -----GAAAAAGTACTTGTATACGAAATGATGTTTGACGGTCTTCTTATCGAA 180
 QY 62 GlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
 Db 181 GGTTC-----GTACGTATCGAAGAATCAATATGATATG 213
 QY 82 LeuCysAlaAspLeuAenSerLeuHisLeuLeuProTyrSer-----GluGlyAlaVal 99
 Db 214 TTACATCCAGATTAGATACCTTGGTAAATCTCCATGCTGCTGGACAGGTAAGTT 273
 QY 100 LeuAlaIleSerAsnProHisAenPheValThrSerGluProLeuPheCysSerProArg 119
 Db 274 GCACGTTTAAATTTGTGATATATAAAACAGATGGAACACCATTTGAAGGGGATCCTCGT 333
 QY 120 ValIleLeuMetGlnGlnIleGluArgLeuAlaAenLeuLysLeuLysGlyLeu---Phe 138
 Db 334 GCAAACTTAAACCGTGTATTAAAAAGAAATTTAGGCTTCACAGACTTTAACCTTA 393
 QY 139 AlaSerGluLeuGluPheAenLeuPheAenGluThrTyrLysSerAlaSerGlnLysHis 158
 Db 394 GGGCTGAAACAGAAATCTCTTGTGTTAAGTTGGATGAANA----- 435
 QY 159 TrpLysAenLeuLysThrAlaGlnProHisGln-----TrpMet 172
 Db 436 -----GGGGAACCAACTTTAGAACTTAATGATGATGGTGATATTTC 477
 QY 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAenLysLeuGlu 192
 Db 478 GATTTAGCACCCTACAGATTTAGTGAAACCTGCTCGTGATATTGTTTAGAATTAGAG 537
 QY 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
 Db 538 GATATGGGCTTCGATATTGAAGCTAGTCACCATGAAGTTGCCCTGGTCAACATGAAT 597
 QY 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
 Db 598 GACTTTAAATATGCAGATGCTGTACAGCATGTGATAATATCCAAACATTTAAATTTGGTT 657
 QY 233 ValArgGluMetAlaGluGlnSerClyMetValAlaThrPheMetAlaLysLeuSerSer 252
 Db 658 GTTAAAAACAATCGCACGTAACATAATTTACACGCAACATTTATGCCCTAAACCATTTTC 717
 QY 253 ThrAlaLeuGlyAenAlaCysHisIleHisMetSerLeuGlnAenAlaGluThrGluLys 272
 Db 718 GGTGTGAATGTAGCGGTATGCACCTTTAACGTTTCATTA-----TTCAAGGTAAGAA 771
 QY 273 AsnAlaPheTyrAspGlnAenAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIle 292
 Db 772 AATGCATTTCTTTGATCCAAATACTCAAAATGGCTTAAACGGAACCTGCATATCAATTTACA 831
 QY 293 AlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSer 312
 Db 832 GCAGGTGTGCTTAAATAATGACGGGATTTGCTGTATGTAAACCCGTTAGTAAACTCA 891
 QY 313 TyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsn 332

Db 892 TATAAAGCTTTAGTACCTGGTTTGAAGCACCATGTTATATGTCATGGAGTGGTAAAC 951
Qy 333 ArgThrSerAlaPheArgLeuGlyCysAsnSerGluGlyIleAsnValGluLeuArg 352
Db 952 CGTTCACCAATTAATCCGTTGACCACTTCAAGAGTATCTACTCGTATCGAAGTACGT 1011
Qy 353 IleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAlaGlyIle 372
Db 1012 TCAGTAGATCCAGCTGCAACCCATACATGCGTGTAGTCAATCTTAGAGCTGGACTA 1071
Qy 373 SerGlyIleGluGlyLeuLeuLeuProProAlaSerGlyAsnValTyr----- 390
Db 1072 GATCGTATTAAATAAATAAATAAAGTTCCAGAACAGTTTACCAAAATATTACGAATG 1131
Qy 391 AsnAspLysGlu-----LeuProGluPheProAsnSerLeuGlnAsnAla 405
Db 1132 AACCGTAGAAGACGTGAAGCAGTAGGACATCAAGACTTACCTTCAACACTTATACGTCA 1191
Qy 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeu 425
Db 1192 TTAAGACGATGCGTGAAGTGAAGTTATTAAAAAGCTTTAGGAATCATATCTATAAT 1251
Qy 426 HisTyrValAsnAlaAlaAsnValGluLeuLeuAsnGluPheSerLysGlnValThrAspTyr 445
Db 1252 CAATTTATTATTCAAAATCAATTTGAATGGATTACTACAGAACTCAAGTCTCTGAATG 1311
Qy 446 GluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1312 GAAAGAGATCATGATCAAGCAATAT 1338

RESULT 13

US-09-815-242-6417
; Sequence 6417, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6417
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1356)
US-09-815-242-6417

Alignment Scores:
Pred. No.: 2,26e-41 Length: 1356
Score: 402.00 Matches: 114
Percent Similarity: 44.81% Conservative: 63
Best Local Similarity: 28.86% Mismatches: 184
Query Match: 16.92% Indels: 34
DB: 9 Gaps: 7
US-10-098-602A-2 (1-454) x US-09-815-242-6417 (1-1356)

Qy 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGly 97
Db 217 AGTGACATGATTATTATACCTCTATCTATCTACATGATGATTTTTCGTCGGAGAACGCAC 276
Qy 98 -----AlaValLeuAlaIleSerAsnProHisAsnPheValThrSer 111
Db 277 CATGGAAAAGTTGCGCGCTTAATTTGGCATATTATACAACTCT-----GATGGT 324
Qy 112 GluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsn 131
Db 325 ACCCTTTTCCGCGAGATCCCGC-----GATTAAT 354
Qy 132 LeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn----- 148
Db 355 TTAAGACGTCCTTAGCAGATATCAAGAAATTAGGTTTACTCTTTTAACTTAGGACCA 414
Qy 149 -----GluThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGln 166
Db 415 GAACCGAAGATTTTCTCTATTAAATTAGATGAGAAAGTGGTGAATATTACCCTGATTGAAT 474
Qy 167 ProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSer 186
Db 475 GATAAAGTGGCTATTGTTGATTTTGGCCCAACCGATTAGTGAAAATTTGTCGCGTGTAT 534
Qy 187 ValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeu 206
Db 535 ATTGTTTATAGATTAGAAAGTTTAGGCTTTGAAGTGAAGCTTCGCATCATGAAGTAGCA 594
Qy 207 ProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHis 226
Db 595 CTTGGTCAACATGAAATTTGACTTTAAATATGCGGATGTAGTTGATGCTGTGCAATATT 654
Qy 227 IleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPhe 246
Db 555 CAAACGTTTAAGTTAGTTGTAATACTATTGCGCAACACATGTTTGCATGCAACATTT 714
Qy 247 MetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGln 266
Db 715 ATGCCAAAACCTTTGTTGGAAATTAATGTTCTGTTATGCTGCACTGCAATATGTCAATTA 771
Qy 267 AspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThr 286
Db 772 ---TTTAATGAAGAAGGTAATGTTTTTATGACGAGTCAGAGAGATGGGCTTTAAGCCAA 828
Qy 287 LeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePhe 306
Db 829 ACTGCTTATCATTTCTTAGTGGTTTATTAAACATGCCCGTCCTATATCTCGGTATGT 888
Qy 307 AlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCys 326
Db 889 AACCCCAACAGTGAACCTTTATAAACGCTTGGTACCAGGTTTATGAAGCACCTGTTTATGTA 948
Qy 327 CysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGly 346
Db 949 GCTTGGAGCGGCAGAAATCGTTCACCACTCATTCGAGTTTCCTGAATTCGTGGTTTATCG 1008
Qy 347 IleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAla 366
Db 1009 ACTCGTTTAGAATTACGTTCTGTCGATCCCTTACGCAATCCGTTATTAAACAATGGCAGTC 1068
Qy 367 IleIleAlaAlaGlyIleSerGlyIleGluGlyLeuLeuLeuProProAlaSer 386

```

1069 TTATTACAAGCTGTTGGATGATTCGTAATGAACCTTACGCCACCGCAGCGATTGAT 1128
QY 387 GlyAsnValTyr-----AsnAspLysGlu-----LeuProGluPhePro 399
Db 1129 CGTAATATTATGATGATGAGGAGGAAACGTCACATGCAAAATGGAAGATTGGCCA 1188
QY 400 AsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPhe 419
Db 1189 TCACCTTACACATGCCATCAAGAAATACGTAAGATCAAGTAATGATTGATGCAATTA 1248
QY 420 GlyGluLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuLeuPheSer 439
Db 1249 GGTGCGTCATATTTTGGCAATTTTGTAGAGCGAAGCAAGTGAATGGCAGCTTCCGT 1308
QY 440 LysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1309 CAAACCGTTTCTGATGGAAGAGAACAAATATTTGGAGTTGTGAC 1353

RESULT 14
US-09-070-927A-188
; Sequence 188, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4176 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 188:
US-09-070-927A-188

Alignment Scores:
Pred. No.: 1.57e-40 Length: 4176
Score: 402.00 Matches: 114
Percent Similarity: 44.81% Conservative: 63
Best Local Similarity: 28.86% Mismatches: 184
Query Match: 16.92% Indels: 34

```

```

DB: 10 Gaps: 7
US-10-098-602A-2 (1-454) x US-09-070-927A-188 (1-4176)
QY 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGly 97
Db 308 AGTGACATGATATTATACCTGACCTATCTACATGATGATTTTCCGTTGGGAAAGCGAC 367
QY 98 -----AlaValLeuAlaIleSerAsnProHisAsnLeuValThrSer 111
Db 368 CATGAAAGTTGCGCGTTAATTTTGTGATATTTCATCTCT-----GATGGT 415
QY 112 GluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsn 131
Db 416 ACCCTTTTGGCGGAGATCCCGC-----GATAAT 445
QY 132 LeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn----- 148
Db 446 TTAATAACGCGCTTAGCAGATATGAAGAATTAGGCTTTACTCTTTTAACTTAGGACCA 505
QY 149 -----GluThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGln 166
Db 506 GAACCAAGATTTTCTCTATTTAAATAGATGAGATGGAATTTACCACTGACTGGAAT 565
QY 167 ProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSer 186
Db 566 GATAAAGTGGCTATTTGATTTTGCACCCACCGATTTAGTGAAAATTTGTCGCGTGAT 625
QY 187 ValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeu 206
Db 626 ATTGTTTAGAGTTAGAAAGTTTAGGCTTTGAAGTGAAGCTTCACATCATGAAGTAGCA 685
QY 207 ProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaPheArgHis 226
Db 686 CCTGTCAACATGAATAATGCTATGATGATGATGATGATGATGATGATGATGATGAT 745
QY 227 IleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPhe 246
Db 746 CAACGTTCAATATTAGTTGTGAATACTATTGCGCGTAAACATGGCTTGCATGCACATTT 805
QY 247 MetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGln 266
Db 806 ATGCCAAAACCTTTGTTGGAATTAATGCTTCTGATGATGATGATGATGATGATGATGAT 862
QY 267 AspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThr 286
Db 863 ----TTTAATGAAGAAGGTAATGTTTTTATGACGAGTCAGAGAGATGGGCTTAGGCCAA 919
QY 287 LeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePhe 306
Db 920 ACTGGGTATCATTTCTTAGTGGTTTATTAAACATGCGCGTGCCTATATCTCGCGTATGT 979
QY 307 AlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCys 326
Db 980 AACCCAAACAGTGAACCTCTTATAACGCTTGGTACCAGGTTATGAAGCACCTGTTTATGTA 1039
QY 327 CysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGly 346
Db 1040 GCTTGGAGCGGAGAAATCGTTCCACCATCTATTCGAGTTCTCGAATCTCTCGTGGTATCG 1099
QY 347 IleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAla 366
Db 1100 ACTCGTTTAGAATTAAGTTCTGCTGATCTCTCAGCGAATCCGTTATTAACTAGGCTGTC 1159
QY 367 IleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSer 386
Db 1160 TTATTACAAGCAGGTTTGGATGATTCGTAATGAACCTTACACCCAGCGCAGTTGAT 1219
QY 387 GlyAsnValTyr-----AsnAspLysGlu-----LeuProGluPhePro 399
Db 1220 CGTAACATTTATGATGATGATGAGGAGAAACGTCACATGCAAAATGGAAGATTGGCCA 1279
QY 400 AsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPhe 419

```

Db 1280 TCAACCTTACCAATGCGCATCAAGAAATACGTAAGATCAAGTAATGATGCAATTA 1339
Qy 420 GlyGluLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluLeuAsnGluPheSer 439
Db 1340 GGTGCTCATATTTTGGCAATTTGTAGAAGCGAAACGAATGGAATGGGCGACGCTTCGCT 1399
Qy 440 LysGlnValThrAspTTPGluLeuAsnGlnGlyPheAsnArgTyr 454
Db 1400 CAAACCGTTTCTGAATGGGAAAGAGAACAAATATTTTGGAGTTGTATC 1444

RESULT 15

US-09-974-300-261
; Sequence 261, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berkta, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-261

Alignment Scores:

Pred. No.: 9.5e-41 Length: 1008
Score: 395.50 Matches: 94
Percent Similarity: 51.25% Conservative: 50
Best Local Similarity: 33.45% Mismatches: 128
Query Match: 16.65% Indels: 9
DB: 10 Gaps: 2

US-10-098-602A-2 (1-454) x US-09-974-300-261 (1-1008)

Qy 181 GluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAla 200
Db 171 GAAAACTGCCCGCGGCATCGTACTTGAGCTCGAAGAAATGGGCTTTGAAATCGAAGCG 230
Qy 201 ThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeu 220
Db 231 TCTCACCACGAAGTTGCACCTGGACAGCAGCAATCGACTTCAATACGCCGCGAGCCATC 290
Qy 221 ThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSer 240
Db 291 CGCGCTTGGCATGACATCCAAACGTTCAAGCTCGTTGTTAAACAAATCGCGCGCAACAC 350
Qy 241 GlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHis 260
Db 351 GGCCTGACGCAACATTCATCCAAAACCGTTGTTGCGCGTAAACCGGATCAGGCGATGCAC 410
Qy 261 IleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAsp 280
Db 411 TGCAATCTATCACTC-----TTCCGAAACCGCGCCACGCTTCTTTGACAAAGAGCGCC 464
Qy 281 GluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPro 300
Db 465 GATCTTCAGCTGAGTGAACCGCAAGCAATTCATCGCGGCATCGTCAAGCAGCGCAACA 524
Qy 301 GluAlaThrTyrPheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr 320
Db 525 AGCTTTACAGCTGTGACAAACCCGACTGTCAACTCTTACAAAGCGCCTTTGTTCCAGGCTAT 584

Qy 321 PheAlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeuCys 340
Db 585 GAAGCAGCTTGTCTATGTGGCAGCGCGCAAAACAGAAAGCCGTTTAATCCGCAATCCG 644
Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
Db 645 GTTTCGCCGCGCATCAGCAGCGCGTGAAGTCAAGAGCGTAGACCTTTCTGCCAACC 704
Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
Db 705 TACCTTGCTTGGAGCTACTTCTTGACGAGGACTTGACGGCATCAAAAACAGCTGGAA 764
Qy 381 LeuProProAlaSerGlyAsnValTyr-----AsnAspLys 393
Db 765 GCACCGGCTCCGATCGACCGCAACATCTATGTCATGACAAAAGAGAGCGCTCGAAAC 824
Qy 394 GluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413
Db 825 GGAATCGTCGACCTTCTCTGCGACACTCGCAGCGCTCTCGAAGAAATTCNAATCAACGAA 884
Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433
Db 885 GTCATGTCGTCAAAGCCCTCGGGGACCATCTATTGGAACACTTCGTCGAGCAAGAAATC 944
Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 945 GAATGGGATATGTTCCGACACACAAAGTCCATCCTTTGGGAGCGGAGCAGTATATGTCAG 1004
Qy 454 Tyr 454
Db 1005 TAT 1007

Search completed: December 18, 2003, 07:57:15

Job time : 19019 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 22:53:00 ; Search time 299 Seconds
(without alignments)
670.194 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MTITYDLNLRNGKIDTV.....INFSKQVTDWELNQGNRY 454

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10098602/runat 17122003 151238 28076/app query.fasta_1.647
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdd
-LIST=45 -DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10098602 @CGN 1.1.56 @runat 17122003 151238 28076 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSBBLACK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	799.5	33.6	4403765	3	US-09-103-840A-2
C 2	799.5	33.6	4411529	3	US-09-103-840A-1
C 3	481	20.2	1350	4	US-09-252-991A-1330
C 4	481	20.2	2538	4	US-09-252-991A-1285
C 5	481	20.2	2571	4	US-09-252-991A-1279
C 6	459	19.3	1347	4	US-09-252-991A-10926
C 7	459	19.3	1395	4	US-09-252-991A-10988
C 8	455	19.1	1395	4	US-09-252-991A-11477
C 9	452	19.0	1452	4	US-09-252-991A-11417
C 10	450	18.9	1470	4	US-09-252-991A-11262
C 11	449	18.9	4259	4	US-09-155-183-12
C 12	443	18.6	1425	4	US-09-252-991A-10987

C 13	432	18.2	13425	4	US-08-961-527-151	Sequence 151, App
C 14	409	17.2	1359	4	US-09-134-001C-1336	Sequence 1336, Ap
C 15	401.5	16.9	1359	4	US-09-252-991A-11454	Sequence 11454, A
C 16	398.5	16.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 17	393.5	16.6	1341	3	US-08-724-814-19	Sequence 19, Appl
C 18	389	16.4	1596	4	US-09-252-991A-10119	Sequence 10119, A
C 19	389	16.4	2400	4	US-09-252-991A-10051	Sequence 10051, A
C 20	360	15.2	3114	4	US-09-252-991A-225	Sequence 225, App
C 21	344.5	14.5	1185	4	US-09-252-991A-11261	Sequence 11261, A
C 22	344.5	14.5	1476	4	US-09-252-991A-2443	Sequence 2443, Ap
C 23	331.5	14.0	1443	4	US-09-328-352-2823	Sequence 2823, Ap
C 24	323.5	13.6	1488	4	US-09-252-991A-1197	Sequence 1197, App
C 25	311	13.1	576	4	US-09-252-991A-11196	Sequence 1196, A
C 26	303	12.8	1437	3	US-08-724-814-15	Sequence 15, Appl
C 27	303	12.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 28	303	12.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 29	293.5	12.4	906	4	US-09-252-991A-2239	Sequence 2239, Ap
C 30	287	12.1	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 31	287	12.1	1830121	4	US-09-543-990A-1	Sequence 1, Appli
C 32	280.5	11.8	1416	4	US-09-252-991A-2530	Sequence 2530, Ap
C 33	280.5	11.8	1791	4	US-09-252-991A-2304	Sequence 2304, Ap
C 34	280.5	11.8	1986	4	US-09-252-991A-2389	Sequence 2389, Ap
C 35	272.5	11.5	756	4	US-09-252-991A-216	Sequence 216, App
C 36	267.5	11.3	1392	4	US-09-252-991A-9863	Sequence 9863, Ap
C 37	265.5	11.2	1446	4	US-09-252-991A-16240	Sequence 16240, A
C 38	265.5	11.2	1500	4	US-09-252-991A-16351	Sequence 16351, A
C 39	255.5	10.8	1470	4	US-09-328-352-237	Sequence 237, App
C 40	253.5	10.7	948	4	US-09-252-991A-16139	Sequence 16139, A
C 41	250.5	10.5	669	4	US-09-252-991A-11057	Sequence 11057, A
C 42	249	10.5	429	4	US-09-252-991A-11056	Sequence 11056, A
C 43	241.5	10.2	573	4	US-09-107-532A-959	Sequence 959, App
C 44	207.5	8.7	420	4	US-09-252-991A-11498	Sequence 11498, A
C 45	194.5	8.2	591	4	US-09-252-991A-208	Sequence 208, App

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores: 2.63e-78 Length: 4403765
Pred. No.: 799.50 Matches: 176
Score: 58.02% Conservatives: 88
Best Local Similarity: 38.68% Mismatches: 180
Query Match: 33.65% Indels: 11
DB: 3 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-103-840A-2 (1-4403765)


```
Db 3172673 ATCCCGGAC---CTGGTCTGGCGGACGCGCAGAGTCTCGCGCTCGCGCAGCATT 3172617
Qy 122 LeuMetGlnGlnIleGluAtrGluAlaAsnLeuLysLeuLysGlyLeu-----Phe 138
Db 3172616 CTGGCGGTGAGCTGATCGG-----CTCAAGCGCGCGGACTGTGCGCGCATGTG 3172566
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrlsSerAlaSerGlnLysHis 158
Db 3172565 GCCACCGAGCTGGAGTTTCATCGTTCGACCGCGGTATCCCGAGGATGGCGCCAGCGGG 3172506
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSer 178
Db 3172505 TATCGCGGCTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3172446
Qy 179 GlyIleGluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMet 198
Db 3172445 CGGATGGAGCGGCTGCTCGCGGACATCGGTGGGTATGGCGGTGCGGTCTCGGATTC 3172386
Qy 199 GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp 218
Db 3172385 GAGCGGTCAAGCGGATCAACATGGCGCCAGCAGGAGATCGGTTTCGTTACGACGAG 3172326
Qy 219 ProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGlu 238
Db 3172325 GCGCTGGTCACTCGCGACAACCATCGCATCTACAAGACGCGCGCAAGGAATCGCGCAG 3172266
Qy 239 GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAla 258
Db 3172265 CAGCAGCGCAAGAGCCTTACGTTTCATGCGGAATACGATGAACGCGAA---GGTAATAGC 3172209
Qy 259 CysHisIleHisMetSerLeuGluAspAlaGluThrGluLysAsnAlaPheTyrlsAspGln 278
Db 3172208 TGTCAATCCATGCTCGCTCGCGGCGCAGCATGCTCGCG---GTGTTTCGCGACAGT 3172152
Qy 279 AsnAspGluThrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrls 298
Db 3172151 AACGGCGCGCAGCGCATGCTCGCATGTTCCGACAGTTCGTCGCGCGCGAGTTGCCACG 3172092
Qy 299 ValProGluAlaThrTyrlsPhePheAlaSerTyrlsAsnSerTyrlsValArgLeuGlnPro 318
Db 3172091 TTGCGGAATTCAGCTGTGATGCGCGCGGACCATTAACCTCAAGCGATTTGCCGAT 3172032
Qy 319 LeuThrPheAlaProThrLysCysTrpAlaIleAspAsnArgThrSerAlaPheArg 338
Db 3172031 AGCAGTTTCGCGCGCAGCGCGCTGCTGGGGCTGGACATCGCATCTGCGCGCTCGGG 3171972
Qy 339 LeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeu 358
Db 3171971 GTGGTT---GGCCACGGCGCAACATCGGGTCGAATCGCGGTTCGCGCGGTGATGTC 3171915
Qy 359 AsnProTyrlsLeuAlaPheSerAlaIleAlaGlyIleSerGlyIleGluLys 378
Db 3171914 AACCAGTACCTGGCGGTGGCGGCTCTCATTCGTGAGGGTTGTACGGTATCGAGCGGGC 3171855
Qy 379 LeuGluLeuProProAlaSerGlyAsnValTyrlsAsnAspLysGluLeuProGluPhe 398
Db 3171854 CTTACAGTTCGCGAGCGCTGTGTCGCGCAACGCTTACCAGCGCGGATGTCGACGCGTG 3171795
Qy 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThr 418
Db 3171794 CCGGTATCGTGGCGCGCGCGCGCTGTGTCGAGGATTCGCGGTGTCGCGGAGCGG 3171735
Qy 419 PheGlyGluLysLeuIleLeuHisTyrlsValAsnAlaAlaAsnValGluIleAsnGluPhe 438
Db 3171734 TTCGCGGAGGATGTGTGTCGCGCATCTACTGAACACGCGGTGTGGAGCTGGCGCGCTTC 3171675
Qy 439 SerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db 3171674 AACCGCGCGGTACCGATTGGGAGAGGATACCTGGATTGAGCGC 3171630
```

RESULT 3

US-09-252-991A-1330

```
; Sequence 1330, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1330
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1330

Alignment Scores:
Pred. No.: 1-35e-46 Length: 1350
Score: 481.00 Matches: 131
Percent Similarity: 50.45% Conservative: 91
Best Local Similarity: 29.77% Mismatches: 185
Query Match: 20.24% Indels: 34
DB: 4 Gaps: 10

US-10-098-602a-2 (1-454) x US-09-252-991A-1330 (1-1350)

Qy 22 LeuAlaCysVal-----AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArg 39
Db 61 ATCGAATGCTGTCACCCCGACCTCAACGCGGTGCGCGCGCAAGGTGATGACCGCGGAG 120
Qy 40 HisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValThrAlaValThr 59
Db 121 GGCTTC-----CTCAGCGCGCGCGCTCGCATCGCCGCT-----GGCGTGTG 165
Qy 60 IleGluGlyIleAlaGlyGlyGlyTyrlsSerValAspThrGlyTyrlsSerAsp 79
Db 166 CTGAATGATCATGCGCGGCTTACCCCGCGCGCTTCTATGCGACGCGAGTGGCGAC 225
Qy 80 CysHisIleCysAlaSerLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaVal 99
Db 226 CTGCGCGTGTGCGCGAGCGCCAGCTCCATCGCTGCGCTGCGCTGGAGC----- 273
Qy 100 LeuAlaIleSerAsnProHisAsnPheValThrSerGlu----- 112
Db 274 -----AACACTCGCGCGCTTTCGCATCTGCGATCGCGAGGAACCTGGATGGCAGC 324
Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnIleGluArgLeuAlaAsnLeu 132
Db 325 CCGTCCGGGTGTCCACCCCGCGCTGCTCAAGAGTGTGGCGCTACGCGCTACGCGCCAC 384
Qy 133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThr 150
Db 385 GGCTGGCAACCGGTGTGGCGCCAGCAACTGGAGTCTTCGCTTCGCGCCCGAACACCGAT 444
Qy 151 Tyrls-SerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHisGln 170
Db 445 CCCAACAGAGCGGTTCAGCGCGCGCTGGCGCGCGATGGCGCGCGAGCTG-GGCTATTC 503
Qy 170 nTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
Db 504 GGCGCTTCAGCGGTGTTCACCAACAGCGCTGCGCGCGCTTCTTCGAGGACGCTTACCGCTG 563
Qy 190 sLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 564 CATGGATGCGCTGGCGCTGTGTGCGGACACCTTCATGACGAGATGGCGCGACCGAGT 623
Qy 210 sGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLys 230
Db 624 CGAGATCAACTCTCTGTCATGGCGACCGCGTCTGCTGCGCGGACGACGACCTTCCTGTTCAA 683
```


Db		2274	CGTCCGGGCTGTCCACCCCGGCCTGCTCAAGCAGGTGGTGCGCGCTTACGCCGCCAC	2215
Qy	133	LysLeuLysGlyLeuPheXalaSerGluLeuGluPheAsnLeuPhe-----AenGluThr	150	
Db		2214	GGCTGGCAACCGGTGGTCGGCCACCGAACTGGATTTCTTCGTCTTCGCCCGCGAACCCGAT	2155
Qy	151	TyrLys-SerAlaSerGlnLysHisTrpLysAenLeuLysThrAlaGlnProHisGln	170	
Db		2154	CCCCACGAGCGGTTCAGGCCCGCGTGGGCGCGATGGCGCGCAGCTG-GGTATTTC	2096
Qy	170	nTrpMetAsnIleSerAlaSerSerGlyLeuGluThrPheMetArgSerValArgAsnLy	190	
Db		2095	GGCTTCAGCGGTGCTTCCAACAACGCGCTCGCGCGCTTCTTCGAGGACGCTTACC	2036
Qy	190	sLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHi	210	
Db		2035	CATGGATGCCTGGCGCTGTGTGGCGCACACCTTCATGCACGAGATGGGCACCGACCAT	1976

QY 210 8G1U6WASBPHNEVA1PFOA1AASBPFOLEU1H1METATAASBPARGH181E1E1EALY 230

1975	CGAGATCAACTTCTCTGCATGGCGACCGGTCTGCTGGCGGCACCAACGCTTCTGTTCAA	1916
Db		---
Qy	230 sHisGlyValArgGluMetAlaGluInSerGlyMetValAlaThr-PheMetAlaLysLeu	250
Db		
Qy	1915 GCACCTGCTCAAGGAGTGGCGTCAAGCAGCGCTGATCGTGGTGTGCATGCCCAAGCC	1956
Db		
Qy	250 uSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluTh	270
Db		
Qy	1855 GCTGGCGAAGACGCCGGCGAGCTCGATGCATCCACAGAGCATCGTGGAACTGGACGG	1796
Db		
Qy	270 rGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAs	290
Db		
Qy	1795 CGGGCGAAACATCTTCAGCGAAGCCGATGGCAG-----CCTTCGGCGCGTTCGC	1742
Db		
Qy	290 nTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrI	310
Db		
Qy	1741 CTTTCATCGGCGGCCACAGGCGTGGCTGGCGGACTTCAGCCCTCTCTCGCGCGCATGT	1682
Db		
Qy	310 aEnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaI	330
Db		

db 1681 GAAC TCCTTCCAGCGCCCTCTGCCATCCCTACGCTTCGCCGAACAACGCCCTGCTGGTCCCA 1622

Qy	330	eAspAsnArgThrSerAlaPheA	ArgLeuCysAsnSerLysSerGluGlyIleAsnValGI	350
		:::	:::	
Db	1621	CGACACCGCGCGCGCTCGGCATCCGCGCAGGAGCCTCCGCGCGCGCGCTCGA	1562	
		:::	:::	
Qy	350	uLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAl	370	
		:::	:::	

DB 1561 GAATCGCC TGGCGGCCGACGCCAACCCCTTACCTGGCCATCGCCGCCAGCCTGGCCGC 1502

[illegible]

Oy		407	sLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTy	427
Db		1393	GCGCTTAAAGGCAGCGCCTGGCGCGAATGTTCGGCAGGAGTTCTCGAAGCTA	1334
Oy		427	rValAsnAlaLaAsnValGLuILeAsnCluPheserLysGlnValThrAspTrgLu	446
Db		1333	CGTGCGCCACAGAGCATGGAGCTGACCAGCTTCTTCGACGAGATCAGGCCCTGGGAG	1276

RESULT 6
US-09-252-991A-10926
; Sequence 10926; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOC

; GENERAL INFORMATION: Rubenfield et al.
 ; APPLICANT: Marc J.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 10926
 LENGTH: 1347
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10926

Alignment Scores:
 Pred. No.: 5,2e-44 Length: 1347
 Score: 459.00 Matches: 126
 Percent Similarity: 46.95% Conservative: 82
 Best Local Similarity: 28.44% Mismatches: 213
 Query Match: 19.32% Indels: 22
 DB: 4 Gaps: 10

US-10-098-602A-2 (1-454) x US-09-252-991A-10926 (1-1347)

QY 16 LysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArg 35
 DB 29 GAGTCCAGTTCGTGACCTCTTATTCGACATGATGAGTGGTGTTCGGGCAAGCG 88
 QY 36 LeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheVal 55
 DB 89 ATCGAAGCAACAGTCTGAACAAAGTCTTCGAAAAGGTATCAATCTCCCGCTTCCCTG 148
 QY 56 TyrAlaValThrIleGluGly-----IleAlaGlyGlyThrGluIleSer 71
 DB 149 TTGCGCTCGATATCACCGCTCCACCGTAGAGAGTACCGCTGCGTCTCGACATCGCG 208
 QY 72 SerValAspThrGlyTyrSerAspCysHisLeuCysAlaAspLeu--AsnSerLeuHis 90
 DB 209 GAGCGCGAC-----CGATCTGCTACCGATCCCGCGACCTCTCC 250
 QY 91 LeuLeuProTrpSerGluGlyAlaValLeuAlaIleSerAsnProHisAsnPheValThr 110
 DB 251 ATGGAACCTTGGCAGAACCGCCGACCGCACTCTGTGACCATGACCACTGGAA 310
 QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
 DB 311 GCGAGCCGTCTCTTCGCCGACCCCGCGAAGTCTCGCCAGTGTGTGCGCAGGTTTACC 370
 QY 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn--Glu 149
 DB 371 GAGATGAAGTACCATCGCTCGCGCTTCGAGCTGGAGTTTACCTGATCGACCGAG 430
 QY 150 ThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHis 169
 DB 431 AAGGTGAACCGCGCGCGCGCGCTCGCTCGCGATCTCCGCAAGGTCG-----484
 QY 170 GlnTrpMetAsnIleSerAlaSerGlyIleGluThrPheMetArgSerValArgAsn 189
 DB 485 CAGTCGCTGAGTGTACTTCCATCGACGACCTCGACGAATACTGAGTGCCTCCAGGAC 544
 QY 190 LysLeuGluAla-----GlyIleLeuMetGluAlaThrHisProGluPheLeu 206
 DB 545 ATCATCGAGCGCGCGCGCGCGCGATCCCGCGCGACGATCGTCCCGCAATCCGA 604
 QY 207 ProSerGlnHisGluLeuAsnPheValProAla---AspProLeuThrMetAlaAspArg 225
 DB 605 CCGCGCGAGTTCGAGTCAACCTCAACACGCTCAACGCGCTTCAAGGCTCGACAC 664
 QY 226 HisIleIleAlaLysHisGlyValArgGluMetAlaGlnSerGlyMetValAlaThr 245
 DB 665 GCGGTCTGTCAAGCGCTGTGTTCAAGAACATCGCTACGACCAACGACGATGACAC 724

QY 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
 DB 725 TTCATGGCCAGCCCTATCCGGCCAGCGCGGACGAGCTGCACGTGCATATCTCGCTG 784
 QY 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer 285
 DB 785 CTCGAC---AAGCATGGCAACAACATCTTACCAGGAGGATCCGAG-----CAGAAC 835
 QY 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhe 305
 DB 836 GCCGATTGGCCCATCGCATCGCGTGTCTCGAGACCTCGCGGCTTCCATGGCCCTTC 895
 QY 306 PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys 325
 DB 896 CTCCTGCCCGAAGTCACTCTACCGCGCTTCGGTTCGACGTTCTACGTGCCGAACGCG 955
 QY 326 CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlu 345
 DB 956 CCGAGCTGGGCGCTGGCAACCGCACCGCTGCGCTGCCGCGGACCGGACCGCGGAC 1015
 QY 346 GlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSer 365
 DB 1016 GCGGTACCGCTGGAAACACCGCGTGGCGCGCGCGCAACCCCTACTGCTACTGGCA 1075
 QY 366 AlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAla 385
 DB 1076 GCGGTGCTGGCAGCGGCTTCATCAGCGGCTGACCAACAGGTCGAGCGCGCGGATC 1135
 QY 386 SerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAla 405
 DB 1136 GAAGGCAACTCTACGACGAGATGAG---CCGAGCCTGCGCAACAACCTGCGCGACGC 1192
 QY 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeu 425
 DB 1193 CTGCGGAGCTGGAGCAAGGAGATCATGCGGAAGTACATCGACCCGGAAGTACATCGAC 1252
 QY 426 HisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp 445
 DB 1253 ATCTTCGTGCGCTGCAAGGAAGCGAGCTGGAGGATTCGAGCATCTCCGACCTC 1312
 QY 446 GluLeuAsn 448
 DB 1313 GAGTACAAC 1321

RESULT 7

US-09-252-991A-10988
 Sequence 10988, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 10988
 LENGTH: 1395
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-10988

Alignment Scores:
 Pred. No.: 5,5e-44 Length: 1395
 Score: 459.00 Matches: 126
 Percent Similarity: 46.95% Conservative: 82
 Best Local Similarity: 28.44% Mismatches: 213
 Query Match: 19.32% Indels: 22
 DB: 4 Gaps: 10

US-10-098-602A-2 (1-454) x US-09-252-991A-10988 (1-1395)

Qy 16 LysileAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArg 35
Db 82 GAGTCCAGTTCGTCGACCTCTATTGACAGATGAATGCGGTTCGCGGAGCGG 141

Qy 36 LeuThrGlyArgHisPheLeuGlyLeuAspGlnLysIleSerIleSerThrPheVal 55
Db 142 ATCGAACGCAACAGTCTTCAAGAAAGTATCAATCTCCCGCTCCCGT 201

Qy 56 TyrAlaValThrIleGluGly-----IleAlaGlyGlyGlyTyrGluIleSer 71
Db 202 TTCGCCCTCGATATACCGGCTCACCCTAGAGAGTACCGGCTCGGTCTCGACATCGC 261

Qy 72 SerValAspThrGlyTyrSerAspCysHisLeuAlaAspLeu---AsnSerLeuHis 90
Db 262 GACGCCGAC-----CGGATCTGCTACCGGCTCCCGGCTCCGATCGGC 303

Qy 91 LeuLeuProTrpSerGluGlyAlaValLeuAlaIleSerAsnProHisAsnPheValThr 110
Db 304 ATGGAACCTGCGAGAGCGCCGACCGCGCACTGCTGTATGACCATGCAAGTGA 363

Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
Db 364 GCGAGCGGTTCCTCGCGGACCCCGGCAAGTCTCGCGCAAGTGTGGCGGAGTTCCAC 423

Qy 131 AsnLeuLeuLeuGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn---Glu 149
Db 424 GAGATGAACTGACCATGTCGCGGCTTCGAGTGTGAGTCTACCTGATGACGACGAG 483

Qy 150 ThrTyrLysSerAlaSerGlnLysHisTyrLysAsnLeuLysThrAlaGlnProHis 169
Db 484 AACGTGAACGCGCGCGCAGCGCTCGCTCGCGGATCTCCGCAAGCGTCCG----- 537

Qy 170 GlnTrpMetAsnIleSerAlaSerGlyIleGluThrPheMetArgSerValArgAsn 189
Db 538 CAGTCGGTGCAGGTGTACTCCATCGACGACCTCGACGAATACGTGAGTGTGCTCCAGG 597

Qy 190 LysLeuGluGluAla-----GlyIleLeuMetGluAlaThrHisProGluPheLeu 206
Db 598 ATCATGACG 657

Qy 207 ProSerGlnHisGluLeuAsnPheValProAla---AspProLeuThrMetAlaAspArg 225
Db 658 CCGCGCGATTCGAGGTCAACCTCAACACGCTCAACGAGCGGCTCAAGCGCTCGACAC 717

Qy 226 HistIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThr 245
Db 718 GCGGTGCTGCTCAAGCGCTGTCAAGAACATCGCTACGACCAACGAGATGACACCA 777

Qy 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
Db 778 TTCATGGCAAGCCCTATCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837

Qy 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer 285
Db 838 CTCGAC---AAGCATGGCAACAACATCTTACACGAGCGAGGATCCGAG----- 888

Qy 286 ThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhe 305
Db 889 GCCGCATTGCGCCATGCGATCGCGGCTGTCTGAGACCCCTCGCGGCTCCATCGGCTTC 948

Qy 306 PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys 325
Db 949 CTCTGCCCGCAACCTCTACCGCGCTTCGGTTCGCGAGTCTTACGTGCGCAACGCG 1008

Qy 326 CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlu 345
Db 1009 CCGAGCTGGGCGCTGGACACCGGACCGTGGCGCTCGCGGCTCCCGGCGGCGGCGGAC 1068

Qy 346 GlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSer 365
Db 365 ATGGAACCTGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 363

Db 1069 GCGGTACGCTGGAACACACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128

Qy 366 AlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAla 385
Db 1129 GCGGTGCTGCGAGCGCTTCATCAGGCTGACCAAGAGTTCGAGCGCGCGCGCGCG 1188

Qy 386 SerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAla 405
Db 1189 GAAGGCAACTCTTACGAGCAGATGGAG---CCGAGCGCTGCGCAACAACCTGCGGCG 1245

Qy 406 ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeu 425
Db 1246 CTGCGGAGTGCAGCGAAGCGAGATCATCGGAAAGTATCATCGACCGAAGTATCAT 1305

Qy 426 HistYrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp 445
Db 1306 ATCTGCTGCTGCGTCAAGGAAAGCGAGTTCGAGCACTCGATCTCGGACCTC 1365

Qy 446 GluLeuAsn 448
Db 1366 GAGTACAAC 1374

RESULT 8
US-09-252-991A-11477
; Sequence 11477, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11477
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11477

Alignment Scores:
Pred. No.: 1,638-43 Length: 1395
Score: 455.00 Matches: 124
Percent Similarity: 47.40% Conservative: 86
Best Local Similarity: 27.99% Mismatches: 211
Query Match: 19.15% Indels: 22
DB: 4 Gaps: 10

US-10-098-602A-2 (1-454) x US-09-252-991A-11477 (1-1395)

Qy 16 LysileAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArg 35
Db 82 GAGTCCAGTTCGTCGACCTCTATTGACAGATGAATGCGGTTCGCGGAGCGG 141

Qy 36 LeuThrGlyArgHisPheLeuGlyLeuAspGlnLysIleSerIleSerThrPheVal 55
Db 142 ATCGAACGCAACAGTCTTCAAGAAAGTATCAATCTCCCGCTCCCGT 201

Qy 56 TyrAlaValThrIleGluGly-----IleAlaGlyGlyGlyTyrGluIleSer 71
Db 202 TTCGCCCTCGATATACCGGCTCACCCTAGAGAGTACCGGCTCGGTCTCGACATCGC 261

Qy 72 SerValAspThrGlyTyrSerAspCysHisLeuAlaAspLeu---AsnSerLeuHis 90
Db 262 GACGCCGAC-----CGGATCTGCTACCGGCTCCCGGCTCCGATCGGC 303

Qy 91 LeuLeuProTrpSerGluGlyAlaValLeuAlaIleSerAsnProHisAsnPheValThr 110
Db 304 ATGGAACCTGCGAGAGCGCCGACCGCGCACTGCTGTATGACCATGCAAGTGA 363

QY	111	SerGluProLeuPheCysSerProArgValIleLeuMetGlnIleGluArgLeuAla	130
DB	364	GGTGATCCCTTCTTCGCGGACCCCGCGAAGTCTCGCCAGGTGGCGCATTCACC	423
QY	131	AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn--Glu	149
DB	424	GAGATCGAATTCACCATCGTCGCGGCTTCGAGCTGGAGTTTCACTGTATCGCAGGAC	483
QY	150	ThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisHis	169
DB	484	AAGCTGAACGGCGCGCGAGCGCGCTCGCCGATCTCCGGCAAGCGTCG-----	537
QY	170	GlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValaArgAsn	189
DB	538	CAGTCGGTCAGGTACTCATCGACGACCTCGACGAATACGTCGAGTGCCTCCAGGAC	597
QY	190	LysLeuGluCluAla-----GlyIleLeuMetCluAlaThrHisProGluPheLeu	206
DB	598	ATCATCGACGGCGCGCGCCAGGGGATCCCGCGGACGCCATCGTCGCCGAATCCGCA	657
QY	207	ProSerGlnHisGluLeuAsnPhe--ValProAlaAspProLeuThrMetAlaAspArg	225
DB	658	CCGCGCAGTTCAGGTCAACTGCATCATCGTCGCGACCCGATGAAGGCTCGACTAT	717
QY	226	HisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThr	245
DB	718	GCGGTGCTGTCAAGGCGCTGATCAAGAACATCGCTTACGACCAACGAGATGGACACC	777
QY	246	PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu	265
DB	778	TTCATGCCAAGCCCTATCCGGGCGAGGAGGAAAGCATGTCAGCTACACATCTCGCTG	837
QY	266	GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSer	285
DB	838	CTCGAC--AAGCAGCGCAACAATCTTCACCAGCGAGGATCCGAG-----CAGAAC	888
QY	286	ThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhe	305
DB	889	GCGCGCATTCGCCATTCGATCGCGCGGTCTCGAGACCTCGCGGCTCCATGGCCCTC	948
QY	306	PheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLys	325
DB	949	CTCTGCCGGAAGCTCAACTCTACCGCGCTTCGGTTCGCGAGTCTACGTGCCGAACGCG	1008
QY	326	CysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGlu	345
DB	1009	CCGAGCTGGGGCTTGGACAACCGCACCGTGGCTCGCGTCCGCCACCGCAGCGCCGAC	1068
QY	346	GlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSer	365
DB	1069	GCGGTACGCTGGAACATCGCTCGCGGCGCGACGCCAACCCCTACCTGCTGCTGCC	1128
QY	366	AlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProProAla	385
DB	1129	TCGGTCTGCCGGGTGCACCGGCTGACCAAGTTCGAGCGGGTGCCCGATC 1188	
QY	386	SerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAla	405
DB	1189	GAAGGCAACTCTACGAGCAGTGGAG--CCGAGCCTGCCGAACACCTGCGCGACGCC	1245
QY	406	ThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeu	425
DB	1246	CTCGCGAATCGACGACGCGAGATCTCGCGAAGTACATCGATCCGAAGTACATCGAC	1305
QY	426	HisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrp	445
DB	1306	ATCTTCGTCGCTCGCAAGGAACGAGCTGGAGGAGTTCGAGTACTCGATCTCGCACTC	1365
QY	446	GluLeuAsn 448	
DB	1366	GAGTACAAC 1374	

RESULT 9

US-09-252-991A-11417/c
; Sequence 11417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11417
; LENGTH: 1452
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11417

Alignment Scores:	3.91e-43	Length: 1452
Pred. No.:	452.00	Matches: 122
Score:	Percent Similarity: 47.58%	Conservative: 84
Best Local Similarity:	28.18%	Mismatches: 205
Query Match:	19.02%	Indels: 22
DB:	4	Gaps: 10

US-10-098-602A-2 (1-454) x US-09-252-991A-11417 (1-1452)

Qy	26	AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp	45
Db	1449	GACATGAATGGCTGGTTCGCGCAAGCAATCGAACCAACAGTCAGTCCCAAGTCTTC	1390
Qy	46	GlnLysLysIleSerIleSerThrPheValTyAlaValThrIleGluGly	62
Db	1389	GA AAAAGGATATCAACCTTCGCGTTCCTCTTCGCGCTCGATATCACCGGCTCCACCGTA	1330
Qy	63	--IleAlaGlyGlyGlyTyArgLysSerValAspThrGlyTyrSerAspCysHis	81
Db	1329	GAGAGTACCGGCTGGTCTCGACATCGGCGACCGCAC-----CGG	1288
Qy	82	LeuCysAlaAspLeu---AsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeu	100
Db	1287	ATCTGTACCCGATCCCGCACCTCTCCATGGAACCTGGCAGAGCGCGACCGCG	1258
Qy	101	AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal	120
Db	1227	CAACTGCTGATACCATGATGAGCTGGAAGGTGATCCCTCTTCGCGCACCCCGCGAA	1168
Qy	121	IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuGlyLeuPheAlaSer	140
Db	1167	GTCTCGCCAGGTGGCGCGGTTTCCAGAGATGGAACTGACCTGTCGCGCGCTTC	1108
Qy	141	GluLeuGluPheAsnLeuPheAsn---GluThrTyrLysSerAlaSerGlnLysHisTyr	159
Db	1107	GAGCTGAGTTCTACCTGATCGACGAGGAAAGCTGACCGCGCGCGCGCGCGCGCG	1048
Qy	160	LysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerGly	179
Db	1047	TCGCGGATCTCCGCAAGCGTCG-----CAGTCGGTGCAGGTGTACTCCATCGACAC	994
Qy	180	IleGluThrPheMetArgSerValArgAsnLysLeuGluGluAla-----GlyIle	196
Db	993	CTGACGAATACGTCGAGTGCTCCAGGACATCATCGACGCGCGCGCGCGCGCGCATC	934
Qy	197	LeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPhe---Val	215
Db	933	CCGCGCACGCCATCTGTCGCGGAATCCGACACCGCGCGAGTTCGAGGTCAACCTCATCAC	874
Qy	216	ProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyValArgGlu	235

```

Db      873  GTCCGCGGACCGATGAAGCGCTGGCACTATCGGTGCTGCTCAAGCGCTGATCAAGAAC 814
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236  MetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeu 255
Db      813  ATCGCTACGACCGAGATGGACACACCTTCATGGCCAAGCCCTATCCGGGCGAGCA 754
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256  GlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPhe 275
Db      753  GGGAAACGAGCTGCAGTACATCTCGTGTGAC---AAGCAGCGGCAACAACATCTTC 697
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
276  TyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeu 295
Db      696  ACCAGCGAGGATCCCGAG-----CAGAACCGCCGATTCGCCCATCGCGGGCGTG 643
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296  LeuLysTrpValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArg 315
Db      642  CTCGAGACCTGCGCGCTCCATGCGCTTCTCTGCCGAAGCTCACTCTTACCGCGC 583
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316  LeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSer 335
Db      582  TTCGGTTCCGAGTTCTACGTGCGAAACGCGCGAGCTGGGGCTTGGACCAACCGCACCGTG 523
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336  AlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGly 355
Db      522  GCCCTGCGCGTGGCCACCGCGAGCCGCGAGCTGAGCTGGAACATCGCGTGGCGCGC 463
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356  AlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSerGlyIle 375
Db      462  GCCAGCGCAACCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376  GluGluLysLeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeu 395
Db      402  ACCAAAGGTGCGAGCGGGTGGCCCGATCGAAGGCACTCTTACGAGCAGTTGGAG--- 346
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
396  ProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeu 415
Db      345  CCGAGCCTGCGGAAACACTCGCGGCGCGCTGCGGCAACTGGAGCAGCAGCAGATCCTG 286
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416  AsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIle 435
Db      285  CGAAGTACATCGATCCGAAGTACATCGACATCTTCGTGCTGCTGCTGCTGCTGCTGCTG 226
Qy      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436  AsnGluPheSerLysGlnValThrAspTrpGluLeuAsn 448
Db      225  GAGGAGTTCGAGTACTCGATCTCGGACTCGAGTACAAAC 187

```

RESULT 10

```

US-09-252-991A-11262
; Sequence 11262, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11262
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11262

```

Alignment Scores:

```

Pred. No.: 6,86e-43 Length: 1470
Score: 450.00 Matches: 130
Percent Similarity: 45.85% Conservative: 80

```

```

Best Local Similarity: 28.38% Mismatches: 222
Query Match: 18.94% Indels: 26
DB: 4 Gaps: 7

```

US-10-098-602A-2 (1-454) x US-09-252-991A-11262 (1-1470)

```

Qy      1  MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db      112  ATGACTACCAAGTTAGACCAAGTGCAGCTGGTGAAGGAACCAAGATCACCGAAGTG 171
Qy      21  ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db      172  GAATGCTGATTTCCGATCTGACAGGCATCGCCGCGCAAGATCTCGCGCAACCAAAA 231
Qy      41  PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db      232  TTCATCCGGAAGGCGCATCGCTGCCGAGAGCGTTCTGCTGCAGACCGTCAACCGC 291
Qy      61  GluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80
Db      292  GACTTACGTGAAGACGACATTTTAC---TAGCAGCTGCTCGACCGCGCGACATCGATG 348
Qy      81  HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSer---GluGlyAlaVal 99
Db      349  GTCTGCGCGCCGACGAGAAATGCGGTTCCTCGTCCCTGGGCGCATCGAGCCACCGCG 408
Qy      100  LeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArg 119
Db      409  ATGGTGTATCCACACACCTTCGACAGCTC---GGCAATCCCATCGAGCTGTCCCGCGC 465
Qy      120  ValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLysGlyLeuPheAla 139
Db      466  AACATCTCAAGCGCGGTGAAGATGATATGCGATTAAGGCTGGCGCGCGATCGTGCG 525
Qy      140  SerGluLeuGluPheAsnLeuPheAsnGluThr----- 150
Db      526  CCAGAGATGAGGTTCCTACCTGACCAAGCGCAGCAGACCCCGACTATCGTTGCGAGCC 585
Qy      151  -----TyrLysSerAlaSerGlnLysHisTyrLysAsnLeuLysThrAlaGlnProHis 168
Db      586  CCGGTGGCGCTCCGAGCGCCAG-----GAAACCGGG 618
Qy      169  HisGlnTrpMetAsnIleSerAlaSerGlyIleGluThrPheMetArgSerValArg 188
Db      619  CGTCAGTCGTCTCCATCGACCGCGCCACAGAGTTCGACCCGCTGTTTCGAGGACATGAC 678
Qy      189  AsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSer 208
Db      679  GACTGTGCGAGCGCCCGAGCGCTTGACCTGATCGATCCACGAGGAAGCGACCGCG 738
Qy      209  GlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIle 228
Db      739  CAGATGAGATCACTTCCTGCTATGGCGAGCGCTCGACCTGCGCGACCAAGATCTCGTG 798
Qy      229  AlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAla 248
Db      799  TTCAGCGCCACCATCGCGAGCGCGCTCAAGCACACAACGCTCGCGCGACCTTCATGCC 858
Qy      249  LysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAla 268
Db      859  AAGCCGATGACCGCGCGAGCGCGCGATGACCTGACCGAGACATCGTCGAGCTG 918
Qy      269  GluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAla 288
Db      919  AAGACCGGCAAGAACATCTTC-----TCCAAATGCGCGAGCGCCATGAGGAGCTGTC 972
Qy      289  ArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrThrPhePheAlaSer 308
Db      973  CTCGACCATATCGCGCGCTGCAGAGATTCATCCCGGAAGTGTGCGCGCTGTTCCGCGCC 1032
Qy      309  TyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrTrp 328
Db      1033  AACGTGAACCTCGTTCCGCGCGCTTCCTCGCGCGATACCTCGCGCGCGGTGAACGTCGAG 1092

```

QY 329 AlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerIysSerGluGlyIleAsn 348
Db 1093 GCGGAAGAAGAACCGTACCGCTGGCGCTGGCGGACTCAGCCCGGAGAACCCCGG 1152
QY 349 ValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIle 368
Db 1153 GTCGAGAACCGTCTCGCCGCGCGGCGGCGGCAACCCCTTACCTGGCGTGGCTGCCAGCC 1212
QY 369 AlaAlaGlyIleSerGlyIleGluGluLeuLysLeuGluLeuProProAlaSerGlyAsn 388
Db 1213 CTGTGCGGTACATCGCATGCTGCGAAGGATCATCAAGCGAGCGCCCGAGGTCAAAGTCC 1272
QY 389 ValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeu 408
Db 1273 GCGTATGAAGCGCGCAACCTG---CGCTGCGCTGACGATCGAAGCGCGCTGGAGCGC 1329
QY 409 LeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrVal 428
Db 1330 ATGGAGAACTGCAAGCCCTCGAGCAGTATCTCGGCGAGCAAGTTCATCAGCGGCTAGCTC 1389
QY 429 AsnAlaAlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTrpGlu 446
Db 1390 GCGGTGAAGCGTGGCGAGCAGCAGAAATTCAGAGCGGTAACTCAGCTCCTGGGAG 1443

RESULT 11

US-09-155-183-12
; Sequence 12, Application US/09155183
; Patent No. 6323011
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN.
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,183
; EARLIER FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 4259
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-12

Alignment Scores:
Pred. No.: 5,11e-42 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 4 Gaps: 6

US-10-098-602A-2 (1-454) x US-09-155-183-12 (1-4259)

QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyLysIleAspThrVal 20
Db 417 ATGAGTAACACCTCGACCGCTCAGCGATGTTGGTGAAGACCAACAGATCAGAGTTC 476
QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
Db 477 GAATGATGATTGGCGACTTGACCGGATCACCGCGGCAAGATCTCGCCCAACCAACAAG 536
QY 41 PheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIle 60
Db 537 TTCATTGCCGAAAGGATCGCCCTGCCGAGAGTGTGCTGTGAGACAGTACGCGGC 596
QY 61 GluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCys 80

Db 597 GACTATGTCGAAGACGACATCTAT---TACGAACCTGCTCGACCCGCGCGACATCGACATG 653
QY 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeu 100
Db 654 ATCTCCGCCCGCGGACGAGTCTCTCTGTCGTCATGGCCATCGAGCCGCGCGG 713
QY 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 714 CAGGTGATTACGACACCTACGACAGCAGGCAACCCGATCGAGCTGTCGCGCAGCAAC 773
QY 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140
Db 774 GTCTCTCAAGAAAGTCTCTCAAACTATTCCGCAAGAGGCTGGCAGCGCATCTGCGCGCG 833
QY 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLys 160
Db 834 GAAATGGAGTTCCTACCTG---ACCAAGCGCAGTGACGACCCCGGATTAC----- 878
QY 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
Db 879 -----CCATTGCAACCGCGGTGGCGGTTCGGGACGTCGCGAAATCGTGCCTAA 929
QY 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
Db 930 TCGTTCTCTATCGAAGCGCCCAACAAATTGACCGCTGTTCCGAAGACGCTCTACGACTGG 989
QY 191 LeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 990 TCGCAACTGCGAGGAGCTGGATCTCGATACGCTGATCCACGAGACGCGCAGGAGT 1049
QY 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLys 230
Db 1050 GAAATCAACTTCGTCAGCGGCGCGCTGCTCCCTGGCGGACCATCTGCTGTTCAG 1109
QY 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
Db 1110 CGCACCATGCGCGAGCGCGCTCAAGCAGCAACGTTGGCGCCGCTTCATGTCGCAAGCG 1169
QY 251 SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
Db 1170 ATGACCGCGGAGCTTGGCGGCGCCATGCACCTGACACGAGCATCATCGATATCGAGACC 1229
QY 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
Db 1230 GCGAAGAACGCTTTC-----TCCATGAAGACGCGGAGCATGAGCCAGTTGTTCTCTCAAC 1283
QY 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
Db 1284 CACATCGCGCGCTCGAGAAATTCATCCTGAACTGCTGCGCTGTTGCGCGCCCAACGTC 1343
QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIle 330
Db 1344 AACTCGTTTCGCGCGTTCCTGCGGACACATTCGCGCGCGGTGAACGTCGAGTGGGCGGAA 1403
QY 331 AspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu 350
Db 1404 GAAATCCGTACCTGGCGCTGGCGGTGCGGATGCGGCGCTCAAAACCGCTGGTGGAA 1463
QY 351 LeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAla 370
Db 1464 AACCGCTCGCGGTGCGGCGCCCAACCGTACCTGGCGGATTCGCGCGAGCTGCTGTGTC 1523
QY 371 GlyIleSerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyr 390
Db 1524 GGTACATCGGATCGTCAAGGATATCAACCCCAAGCGCGCTGCTGGTGGTTC 1583
QY 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLys 410
Db 1584 GAGCGCGCAACCTG---CGTCTGCGCTGACCATCGAAGACGCTCTGGAACGCGTGGAA 1640
QY 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAla 430

Db 1641 AACAGCAAGACCATCGAGAAATACCTGGGTCAACAATTCATCATCTGGCTAGCTCGCGGTC 1700
 Qy 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446
 Db 1701 AAGCGGCGGAGCATGAAACTTCAAGCGCGTGATCATGCTCATGGAA 1748

RESULT 12

US-09-252-991A-10987/c
 ; Sequence 10987, Application US/09252991A
 ; Patent No. 6551795
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10987
 ; LENGTH: 1425
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-10987

Alignment Scores:

Pred. No.: 4,34e-42 Length: 1425
 Score: 443.00 Matches: 128
 Percent Similarity: 45.92% Conservative: 80
 Best Local Similarity: 28.26% Mismatches: 219
 Query Match: 18.64% Indels: 26
 DB: 4 Gaps: 7

US-10-098-602a-2 (1-454) x US-09-252-991A-10987 (1-1425)

Qy 6 AspGluLeuAsnLeuIleArgAsnGlyLysIleAspThrValValLeuAlaCysVal 25
 Db 1422 GACAGCTGACCATCTGGTGAAGACGACGAGATCCAGGATGAGTGGATGCTGATTTC 1363
 Qy 26 AspMetGlnArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45
 Db 1362 GATCTGACGAGCATCGCGCGGCAAGATCTCGCGGCAACAATAATTCATCGCGGAAAG 1303
 Qy 46 GlnLysLysIleSerIleSerThrPheValThrAlaValThrIleGluGlyIleAlaGly 65
 Db 1302 GGCATGCGCTCGCGGAGAGCGTCTCTGTCAGACCGCTACCGCGGACTAGTGGAAAGAC 1243
 Qy 66 GlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeuCysAlaAsp 85
 Db 1242 GACATTATC---TACGACCTGCTCGACCGCGCGGACATGATATGGTCTGCGCGCGGAC 1186
 Qy 86 LeuAsnSerLeuHisLeuLeuProTrpSer---GluGlyAlaValLeuAlaIleSerAsn 104
 Db 1185 GAGAATCGCGTTCCTCTGCTCGCTCGCCATCGAGCCACCGCGGATGGTGTATCCAGAC 1126
 Qy 105 ProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMetGln 124
 Db 1125 ACCTTCGCAAGCTC---GGCAATCCCATCGAGTGTGCGCGGCAACATCTCTCAAGCGC 1069
 Qy 125 GlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPhe 144
 Db 1068 GTGTGAAGATGATGTCGATAAGGGCTGGCGCGGATGTCGGCGCCAGAGATGGATTC 1009
 Qy 145 AsnLeuPheAsnGluThr-----TyrLysSer 153
 Db 1008 TACCTGACCAAGCGCAGCAGCCCGACTATCCGTTGAGGCCCGCGGTGGCGCGCTCC 949
 Qy 154 AlaSerGlnLysHisIleTrpLysAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsn 173
 Db 948 GGACGCCAG-----GAAACCGGCGCTCAGTCTGTTCTCC 916

Qy 174 IleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGluGlu 193
 Db 915 ATCGAGCGCCCAACAGATTTCGACCGCTGTTTCGAGGACATGTACGACTGTCGAGGCC 856
 Qy 194 AlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsn 213
 Db 855 CAGGGCCCTTGACCTGGATAGCTGATCCACGAGAGGACCGCGGAGATGGAGATCAAC 796
 Qy 214 PheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGlyVal 233
 Db 795 TTTCGTCATGGCAGCGCCCTTCGACCTGGCGCAGATCTCTGTTTCAAGCGCACCATTG 736
 Qy 234 ArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerThr 253
 Db 735 CGCGAGCGCGCTCAAGCACAACGTCGCGCGACCTTCATGGCCCAAGCGGATGACCGGC 676
 Qy 254 AlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsn 273
 Db 675 GAGCGCGGACGCGATGCACTTCACACGACGATCTGTCGACGTGAAGACCGGCAAGAAC 616
 Qy 274 AlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAla 293
 Db 615 ATCTTC-----TCCAAATGCCGACGCGCATCATGAGCGAGCTGTTCTGACCATATCGGC 562
 Qy 294 GlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyr 313
 Db 561 GGCTGCGAGAGTTTCATCCGGAAGTCTCCCGCTGTTTCGCGCCCAACGTGAACCTGTC 502
 Qy 314 LysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArg 333
 Db 501 CGCGCTTCTGCGCGCATCTCGCGCGCGGTGAACGTCGAGTGGGCGGAGAGAACCGT 442
 Qy 334 ThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIle 353
 Db 441 ACCGTGCGCTGCGCGTGGCGGACTCCAGCCCGAGAACCGCGGTCGAGAACCGCTCTC 382
 Qy 354 GlyGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleIleAlaAlaGlyIleSer 373
 Db 381 CGCGCGCGGACGCCAACCTTACCTGGCGCTGGCTGCGGCGGCTGTCGCGCTACATC 322
 Qy 374 GlyIleGluLysLeuLeuLeuProProAlaSerGlyAsnValTyrAsnAspLys 393
 Db 321 GGCATGTGCAAGGCTCAAGCGCGGCGGCTCAAGGTCAAAGGTGCGGCTATGAACGCGC 262
 Qy 394 GluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413
 Db 261 AACCTG---CGCTGCGCTGACGATCGAAGCGCGCTGGAGCGCATGGAGAACTGCAAG 205
 Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433
 Db 204 CCCTTCGACGATATCTCGCAGCAAGTTTCATCAGCGGCTACGTCGCGTGAAGCGTGC 145
 Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446
 Db 144 GAGCAGCAGAAATTTCAAGCGGGTAATCAGCTCTCTGGAG 106

RESULT 13

US-08-961-527-151/c
 ; Sequence 151, Application US/08961527
 ; Patent No. 6420135
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

```

;
; OPERATING SYSTEM:  MSDOS version 6.2
;

```

```
SOFTWARE:  ASCII Text
```

;; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

: TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 151:

SEQUENCE CHARACTERISTICS:

LENGTH: 13425 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

```
! ;  
; TOPLOGY: linear
```

US-08-961-527-151

Alignment Scores:

arrayname	score:	
pred. No.:	3.33e-39	Length: 13425

score:	432.00	Matches:	128
score:	432.00	Matches:	128

Score:	432.00	Matches:	120
Percent similarity:	44.80%	Conservative:	83

Percent Similarity:	44.80%	Conservative:	83
Best Local Similarity:	27.18%	Mismatches:	220

Best Local Similarity:	27.18%	MissMatches:	220
Query Match:	78.18%	Indels:	40

Query Match:	10.10%	Inders:	40
DR:	4	Gang:	10

US-10-098-602A-2 (1-454) x US-08-961-527-151 (1-13425)

Qy	1	MetThrIleThrTyrAspGluLeuAsnLeuIleArgAsnGlyLeuIleAspThrVal	20
Db	11080	ATGCCAAATCGACAGCTTCGTGGTGAAGTCAAGGAAAAAATGTTACCTTTATT	11021
Qy	21	ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLeuArgLeuThrGlyArgHis	40
Db	11020	CGTCTTATGCTTCAGATATTTTGGGAACCATGAAAAACGTCGAANAATTCCT	10970
Qy	41	PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla	57
Db	10969	-----GCTACAGATCAACAGTTAGATTAAGGCTTGTGCAACAAGTTATGTTGATGGA	10916
Qy	58	ValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyr	77
Db	10915	TCTTCTATTGAAGTTT-----GTACGTATCAATGAG	10883
Qy	78	SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGlu---	96
Db	10882	TCGGATATGCTACTGTACCCGACCTGGATACATGCAGCAGTCTTCCTTGGGAGATGAA	10823
Qy	97	---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe	115
Db	10822	AATGGAAAGTGTTCGAGTCTGATCTGTGATGTCTATACACAAAGGTGAACCATTTGCG	10763
Qy	116	CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLeuLys	135
Db	10762	GGTGACCTCGTGGTAATTTGAAACGAGCTCTTCGTCAATGGAAGAAGTTGGATCCAA	10703
Qy	136	GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPhe-----AsnGluThrTyrLys	152
Db	10702	TCCTTCAACCTTGGTCCAGAGCCAGAAATCTTCCTATTTAAGTTGGATGAAAAATGGGAC	10643
Qy	153	SerAlaSerGlnLysHisIleTrpLysAsnLeuLysThrAlaGlnProHisGluTrpMet	172
Db	10642	CCAACACTTGAAGTGAATGACAAAGGT-----GGTACTATT	10607
Qy	173	AsnIleSerAlaSerSerGluIleGluThrPheMetArgSerValArgAsnLysLeuGlu	192

; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1336
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1336

Alignment Scores:
Pred. No.: 4e-38 Length: 1359
Score: 409.00 Matches: 111
Percent Similarity: 47.84% Conservative: 77
Best Local Similarity: 28.24% Mismatches: 175
Query Match: 17.21% Indels: 30
DB: 4 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-134-001C-1336 (1-1359)

Qy 78 SerAspCysHisLeuLeuValAlaAspLeuAsnSerLeuHisLeuLeuProTyrSer----- 95
Db 220 TCAGATATGATATTACATCTGATTTAGATACCTGGGTATCTTCCCTTGGACTGCTGGA 279
Qy 96 GluGlyAlaValLeuAlaHisSerAsnProHisAsnPheValThrSerGluProLeuPhe 115
Db 280 CAAGGAAAGTTGCAGCACTAATCTGTGATGATTTAAACAGATGGTACACCATTTGAA 339
Qy 116 CysSerProArgValLeuMetGlnGlnLeuArgLeuAlaAsnLeuLeuLeuLeuLeu 135
Db 340 GGTATCCACGAGCTAATCTGAGCGTGTATTAGAACGATGGAAGATATGGCTTTACT 399
Qy 136 GlyLeu-----PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLeuSerAla 154
Db 400 GATTTTAATCTAGGCGCTGAACCAAGATTTTCTTATTAAATAGACGAAAAA----- 453
Qy 155 SerGlnLysHisTyrPheAsnLeuLysThrAlaGlnProHisHisGln----- 170
Db 454 -----GGCGAACCTACATAGATTAATTAACAGCAT 483
Qy 171 -----TrpMetAsnIleSerAlaSerSerGlyLeuGluThrPheMetArgSerValArg 188
Db 484 GGTGGTTATTTCGATTTAGCTCCTACAGATTTAGTGAAATTTGCGCGGTGACATCGTT 543
Qy 189 AsnLysLeuGluGluAlaGlyLeuMetGluAlaThrHisProGluPheLeuProSer 208
Db 544 TTAGAATTAGAAGATATGGCTTTGACATTGAAGCAAGCCACCATGAAGTACGCCAGGT 603
Qy 209 GlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIle 228
Db 604 CAACATGAATTTGACTTTTAAATATGCAGATCCGTTACAGCATGTGATAATATCAAAACA 663
Qy 229 AlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAla 248
Db 664 TTTAACTAGTTGTTAAACAACTTGCAGTGAAGCATAATTTACATGCAACATTTATGCCA 723
Qy 249 LysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAla 268
Db 724 AAACCATTTATTTGTTGTAACCGTAGTGTATGCATTTCAACGTATCACTA-----TTT 777
Qy 269 GluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAla 288
Db 778 AAAGGAAAGAGATGCGCTTTCTTGTATCTTGAAGGTGATTTTACAAATTTGACTGATACTGCA 837
Qy 289 ArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPhePheAlaSer 308
Db 838 TATCAATTTACAGCTGGTGTCTTTAAACCGCTAGAGGATTCACCTGCAGTATGTAATCCA 897
Qy 309 TyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTyr 328
Db 898 ATTGTCAACTCATATAAGCTTTGTACACAGTTACGAAGCACCATTGTTATTATTCATGG 957
Qy 329 AlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyLeuAsn 348
Db 958 AGTGGTAAACCCGTTACCTTTTAGTACGTGTCTCAACATCTAGAGGTCTATCAACTCGT 1017

RESULT 15

US-09-252-991A-11454
; Sequence 11454, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11454
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11454

Alignment Scores:

Pred. No.: 3.05e-37 Length: 1359
Score: 401.50 Matches: 103
Percent Similarity: 49.27% Conservative: 65
Best Local Similarity: 30.21% Mismatches: 162
Query Match: 16.90% Indels: 11
DB: 4 Gaps: 7

US-10-098-602A-2 (1-454) x US-09-252-991A-11454 (1-1359)

Qy 113 ProLeuPheCysSerProArgValIleLeuMetGlnGlnLeuGluArgLeuAlaAsnLeu 132
Db 2 CCCTTCTTCGCCAGCCCGCGGAGTCTCGCCAGGTGTGGCCAGGTTTCACCGAGATG 61
Qy 133 LysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsn---GluThrTyr 151
Db 62 GAACGTGACCATGTGCGCGCTTCGAGCTGGAGTTCTACCTGTATCGCAGGAGAAAGCTG 121
Qy 152 LysSerAlaSerGlnLysHisTyrLysAsnLeuLysThrAlaGlnProHisHisGlnTrp 171
Db 122 AACGGCGCGCGAGCGCGCGCTCGCCGATCTCCGGCAAGCGTCCG-----CAGTCG 175
Qy 172 MetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeu 191
Db 176 GTGAGGTGTACTCCATCGACGACCTCGACGAATACGTGCGTGCCTCCAGGACATCATC 235

Search completed: December 18, 2003, 04:28:12
Job time : 7167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 22:45:30 ; Search time 394 Seconds
(without alignments)
3110.520 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MTITYDELNNLIRNGKIDTV.....INBFSKVQTDWELNQGFNRY 454

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cnp2_1/USPTO.spool/US10098602/runat_17122003_151235_28032/app_query.fasta_1.647
-DB=N_Censeq_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCLAMT=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10098602 @CNG 1.1 312 @runat_17122003_151235_28032 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	799.5	33.6	4403765	22	AAI99683	Mycobacterium tube
C 2	799.5	33.6	4411529	22	AAI99682	Mycobacterium tube
C 3	614.5	25.9	1860	22	AAI68686	Pseudomonas sp KIE
C 4	614.5	25.9	11355	22	AAI68692	Pseudomonas putida
C 5	502.5	21.1	34088	23	AAI59566	Propionibacterium
C 6	475.5	20.0	5500	24	ABQ82091	Brevibacterium lac
C 7	470.5	19.8	1338	22	AAH67407	C glutamicum codin
C 8	470.5	19.8	1461	22	AAH67407	Cornebacterium gl
C 9	470.5	19.8	34980	22	AAH68531	C glutamicum codin
C 10	464	19.5	2365589	24	ABA90521	Genomic sequence o
C 11	453	19.1	34980	24	ABQ81848	Bifidobacterium lo
C 12	449	18.9	4259	18	AAV04201	Merged contigs pfi
C 13	445	18.7	1344	24	ABN68493	Streptococcus poly
C 14	445	18.7	2155561	24	ABN71527	Streptococcus poly
C 15	441	18.6	1344	24	ABN68494	Streptococcus poly
C 16	432.5	18.2	1373	22	APF0972	P. putida KT2440-a
C 17	432	18.2	1344	25	ABX06145	S. pneumoniae type
C 18	432	18.2	1347	23	AAI55643	Streptococcus pneu
C 19	432	18.2	13425	19	AAV52284	Streptococcus pneu
C 20	432	18.2	2162598	25	ABS56454	Streptococcus pneu
C 21	426	17.9	1348	25	AAI51270	Streptococcus pneu
C 22	424.5	17.9	1338	23	AAI52096	Staphylococcus aur
C 23	424.5	17.9	1341	23	AAI554603	Staphylococcus aur
C 24	424.5	17.9	1341	23	AAI55040	Staphylococcus aur
C 25	411	17.3	495269	24	ABQ67195	Listeria innocua c
C 26	410	17.3	3011208	24	ABQ69245	Listeria innocua D
C 27	410	17.3	3497	24	ABQ70835	Listeria monocytog
C 28	410	17.3	2944528	24	ABA03041	Listeria monocytog
C 29	409	17.2	1341	22	AAH53428	S. epidermidis ope
C 30	409	17.2	1359	24	ABN91873	Staphylococcus epi
C 31	402	16.9	1356	23	AAI52780	Enterococcus faeca
C 32	402	16.9	4176	20	AAI13125	Enterococcus faeca
C 33	402	16.9	4176	24	ABN98920	Enterococcus faeca
C 34	400.5	16.9	1329	10	AAI92017	DNA encoding pepti
C 35	397.5	16.7	1664976	19	AAV21209	Methanococcus jann
C 36	395.5	16.6	1008	24	ABK72970	Bacillus lichenifo
C 37	393.5	16.6	1341	19	AAV29071	Mycobacterium tube
C 38	393.5	16.6	38110	25	AAI49662	Mycobacterium tube
C 39	393.5	16.6	38110	25	AAI53730	Ribonuclease P RNA
C 40	389.5	16.4	265118	22	AAH41227	Pyrococcus abyssi
C 41	387	16.3	3666	22	AAH54179	S. epidermidis gen
C 42	351	14.8	989	19	AAZ96252	S. pneumoniae deri
C 43	351	14.8	989	19	AAV42971	Streptococcus pneu
C 44	338.5	14.2	3861	22	AAH54760	S. epidermidis gen
C 45	330.5	13.9	1344	22	AAI61030	P. putida KT2440-a

ALIGNMENTS

RESULT 1
AAI99683/C
ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX
OS variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; de.
XX
OS Mycobacterium tuberculosis.
XX
FN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
DR Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ.
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Alignment Scores:
Pred. No.: 4.58e-71 Length: 4411529
Score: 799.50 Matches: 176
Percent Similarity: 58.02% Conservative: 88
Best Local Similarity: 38.88% Mismatches: 180
Query Match: 33.65% Indels: 11
DB: 22 Gaps: 7

US-10-098-602A-2 (1-454) x AA199682 (1-4411529)

Qy 3 IleThrTyrAspGluLeuAsnAsnLeuLeuArgAsnGlyLysIleAspThrValValLeu 22
Db TTGGCGGTGGAGGAGTGGAGCGATGCTGCGCGCGGTGACGTGCGACCGCTCATCGTC 3172914
Qy 23 AlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu 42
Db CGGTTCACCGACATGACAGGCGCGGTGCGCGGAAACGGATATCGGCGCGCATTTGCTC 3172854
Qy 43 --GlyLeuAspGlnLysLysLysIleSerThrPheValTyrAlaValThrIleGlu 61
Db GACGACATAGCCACCGCGGTGCGAGTGTGCGAGTATCTGCTGGCGGTGGAGTGCAC 3172853
Qy 62 GlyIleAlaGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHis 81
Db CTGAACACGGTGGCGGTATGCGATGCGACATTTGGGACACCGGCTACGGCGATGCTG 3172734
Qy 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla 101
Db ATGACGCGGAGTGTCCACTCTCGGCTGATTCCTTGGTACCGGGAACGGCGCTGGTG 3172674

Qy 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121
Db ATGCGCGAC--CTGGTCTGGCGCGAGCGAGCGAGTGGCGGCTCGCGCGAGCATT 3172617
Qy 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeu-----Phe 138
Db CTGCGCGTCTGAGTCTGATCGG-----CTCAAGCGCGCGGAGTGGTTCGCGCATGTG 3172566
Qy 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHis 158
Db GCCACCGAGTGGAGTTCATGCTTCGACCGCGGTATGCCAGGATATGGCCAGCGGG 3172506
Qy 159 TrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSer 178
Db TATCGCGGCTGACCCCGCGCGAGCATCAACATCATGACTACGCGATATTGGCATCTCTCG 3172446
Qy 179 GlyIleGluThrPheMetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMet 198
Db CGGATGAGCGGTGCTGCGCGACATCCGTTGGGTATGGCGGTGCGGCTCTCGCATTC 3172386
Qy 199 GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp 218
Db GAGCGGTCAAGCGGATGCAACATGGCCAGCAGAGATCGGTTTCGTACACGAG 3172326
Qy 219 ProLeuThrMetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGlu 238
Db GCGCTGCTACCTGCGCAACCATGCGATCTACAAGAACGGCGCAAGAAATCGCGAC 3172266
Qy 239 GlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAla 258
Db CAGCAGCGCAAGAGCCCTAACGTTTCATGGCGAAATACGATCAACCGCAA--GGTAATAGC 3172209
Qy 259 CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGln 278
Db TGTCACTCATCTGCTGCTGCGTGGCGACGATGGTCCGCG--GTGTTGCCACAGT 3172152
Qy 279 AsnAspGluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyr 298
Db AACGGCGCGCAGCGCATGCTGCTGTTCCGCGAGCTTCGTCGCGCGCCAGTTGCCACG 3172092
Qy 299 ValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnPro 318
Db TTGCGCGAATTCAGCTGCTGCTATGCGCGCACCATTAACCTCTACAAGCGATTTGCCGAT 3172032
Qy 319 LeuThrPheAlaProThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArg 338
Db AGCAGTTTCGCGCGCGCGCGTGGTGGTGGCGGTGGACAACTGCACCTCGCCCTCGCG 3171972
Qy 339 LeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyIleAlaAspLeu 358
Db GTGGTT--GGCCACGGGCAAAACATCCGGTTCGATGCGCGGTTCGCGCGGTGATGTC 3171915
Qy 359 AsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLys 378
Db AACAGTACCTGGCGGTGGCGGTCTCATTCGTGGAGGTTGTACGATATCGAGCGGGC 3171855
Qy 379 LeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPhe 398
Db CTTGAGTGGCGCGCGCTGTCGCGCAACCGCTTACCAGCGCGCGATGTCGACGCGGTG 3171795
Qy 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLysGluSerLysMetLeuAsnLysThr 418
Db CCGGTTACGCTGGCGCGCGCGGTGCTGTCGAGGATTCGCGGTGGTGGCGGAGCG 3171735
Qy 419 PheGlyLysLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPhe 438
Db TTGCGCGAGATGTGTGTCGCGCACTACCTGAACAACCGCGGTGTGGAGCTGGCGGCTTC 3171675
Qy 439 SerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
Db AACCGCGCGTCCCGATTTGGGAGAGGATACGTGGATTTGAGCGC 3171630

RESULT 3


```

Db 1414 ATCGAAATAAATCAACGCCGAAAGATCCAATCTCTCGGTAAACGCTTAC-----AAGATA 1467
Qy 395 LeuProGluPheProAsn-----SerLeuGlnAsnAlaThrHisLeuLeuLys 410
Db 1468 TCGCGAGAGCTCGTCGCCACTCGCGCGCTCATTTGGAGAGGCTGCGCGCATTTTCGC 1527
Qy 411 GluSerLysMetLeuAsnLysThrPheGlyGluLysLeuLeuLeuHisTyrValAsnAla 430
Db 1528 GAAAGTGAATGGCGGAGTGATCTTCCCTAACGAGTTCGTTGAGCACTATGCCAGATG 1587
Qy 431 AlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTyrGluLeu 447
Db 1588 AAAGTTTGGAGATTAAACAGTCCAATAGTTTCGTGTAATAATGGGAGTTG 1638

RESULT 4
ID AA168692
XX AA168692 standard; DNA; 11355 BP.
AC AA168692;
XX
DT 17-JAN-2002 (first entry)
XX
DE Pseudomonas putida ipuA, B, C, D, E, F, G and H encoding DNA.
XX
KW L-alaninol; isopropylamine; iupI; ipuH; ofloxacin; ipuC; glutamylamide;
KW gamma-glutamylamide synthase; theanine; ipuA; ipuB; ipuD; ipuE; ipuF;
KW ipuG; ipuH; ds.
XX
OS Pseudomonas putida.
XX
FH Key Location/Qualifiers
FT CDS 1314..2340
FT FT /*tag= a
FT FT /product= "ipuA"
FT FT 2342..2680
FT FT /*tag= b
FT FT /product= "ipuB"
FT FT 2743..4122
FT FT /*tag= c
FT FT /product= "ipuC"
FT FT 4194..5354
FT FT /*tag= d
FT FT /product= "ipuD"
FT FT 5371..5565
FT FT /*tag= e
FT FT /product= "ipuE"
FT FT 5589..6476
FT FT /*tag= f
FT FT /product= "ipuF"
FT FT 6533..7963
FT FT /*tag= g
FT FT /product= "ipuG"
FT FT 8051..9574
FT FT /*tag= h
FT FT /product= "ipuH"
FT FT 9600..9698
FT FT /*tag= i
FT FT /note= "Unidentified partial ORF, no start or stop
FT FT codon given"
XX
PN WO200173038-A2.
XX
PD 04-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-EP03651.
XX
PR 31-MAR-2000; 2000EP-0106888.
XX
PA (LONZ ) LONZA AG.
XX
PI Leisinger T, Van Der Ploeg J, Kiener AM, De Azevedo Waesch SI;
PI Maire T;
XX

```

```

DR WPI; 2001-626266/72.
DR P-PSDB; AAG80036, AAG80037, AAG80038, AAG80039, AAG80040, AAG80041,
DR AAG80042, AAG80043, AAG80044.
XX
PT New microorganism for converting isopropylamine to L-alaninol, useful
PT as pharmaceutical intermediate especially ofloxacin, has the ipuH and I
XX genes for alaninol catabolism inactivated
XX
PS Claim 8; Fig 3A-N; 106pp; German.
XX
CC This invention describes novel microorganisms (A), or their enzymatic
CC extracts, able to convert isopropylamine (I) to L-alaninol (II) and which
CC have the ipuH and ipuI genes, encoding enzymes involved in metabolism of
CC (II), inactivated. (II) is an intermediate for pharmaceuticals,
CC especially ofloxacin. Also the enzyme encoded by the bacterial ipuC gene
CC (a gamma-glutamylamide synthase) is used for synthesis of gamma-
CC glutamylamides, specifically theanine. Inactivation of the ipuH and I
CC genes improves production of (II). Typically, a Pseudomonas in which both
CC ipuH and I genes were inactivated produced (II) at 8 mM after 60 hour
CC culture. For a similar strain in which only the ipuI had been inactivated
CC the maximum (II) concentration was about 7 mM, after 6 hour, falling to
CC practically zero after about 20 hour. This sequence encodes the
CC Pseudomonas putida ipuA, ipuB, ipuC, ipuD, ipuE, ipuF, ipuG and ipuH
CC protein products which are described in the method of the invention.
XX
SQ Sequence 11355 BP; 2508 A; 2720 C; 3264 G; 2862 T; 1 other;

```

Alignment Scores:

```

Pred. No.: 5,95e-55 Length: 11355
Score: 614.50 Matches: 159
Percent Similarity: 50.77% Conservative: 73
Best Local Similarity: 34.79% Mismatches: 194
Query Match: 25.86% Indels: 31
DB: 22 Gaps: 10

```

US-10-098-602a-2 (1-454) x AA168692 (1-11355)

```

Qy 7 GluLeuAsnAsnLeuLeuArgAsnGlyLysIleAspThrValValLeuAlaCysValAsp 26
Db 2773 AAGTCCGAGATTTATTGAGAAACACATATCGACAGATTAGCTGGGTGGCAGTCAT 2832
Qy 27 MetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu-----Gly 43
Db 2833 ATAGATGGAGTCTGGCGCGCTAAGCAAGTGGCGCAGAGTATTCTTGAATAAGCGGCT 2892
Qy 44 LeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGluGlyIle 63
Db 2893 CTTGATGGGACGAGATCTCTAAACATTCTCTTTGGATGGGATGTGGCCGATCACTTCGTC 2952
Qy 64 AlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeuCys 83
Db 2953 -----GACGTTTGGAGTTTACTGGCTGGGATTCGGGTATCCGGATATCGCCTTAATT 3006
Qy 84 AlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAlaIleSer 103
Db 3007 CCTGATCTTTCGACTCTTTTCGCTGGTACCCTGGCAGGAAAAAACTGCTTCGCTCTTTCG 3066
Qy 104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123
Db 3067 GATATCCAGCAT--CTAAATGGCGAGCCGCTGAATCTTTTCGCTAGGAACCTTCTCGCT 3123
Qy 124 GlnGlnIleGluArgLeuAlaAsnLeuLysGlyLeuPheAlaSerGluLeuGlu 143
Db 3124 AAAGCAATCGAAAAAGCAGACAGAGCTTGGTTACAGTGCTACGCTGCTTACGAGTTTGAG 3183
Qy 144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisIleTyrLysAsnLeuLys 163
Db 3184 TTTTACCTTCTTAACGACTCTATCGCGAGTATTTCGCGACATCAGTCGCGCAGCATTAAC 3243
Qy 164 ThrAlaGlnPro-----HisHisGlnTrpMetAsnIleSer 175
Db 3244 CCGGTGAAAAAGAGCGGCACCTGCTACGACATGCTCCATCAC-----TCT 3288

```


Qy 74 AspThrGlyTyr-----SerAspCysHisLeuCysAlaAspLeuAsnSer 88
Db 18790 GAG---GGGTATGCCCGCGTCTACGAGCAGATGTGTGGCCACCCCGATCCGCCAC 18734
Qy 89 LeuHisLeuLeuProTsrSerGluGlyAlaValLeuAla-----IleSer 103
Db 18733 TTCAAAGTTTGGCATCGCTTCTGGCACCAGCAGCCAGGATGTTCTCGGATATCACC 18674
Qy 104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123
Db 18673 AATCTCTGACGGA-----ACCTCTCTGCA-----GCTGATCTCTGTCAGTCTCAAG 18626
Qy 124 GlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGlu 143
Db 18625 AGGACGATGGCTGTCGCCGAGAAATGGGTTTACCTTCTATGTTCCACCAGATTCAG 18566
Qy 144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTsrPheAsnLeuLys 163
Db 18565 TTCTACCTGCTTAACGAT-----GAGCATCATCTCTGGGGAATCCCGC 18524
Qy 164 ThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPhe 183
Db 18523 GTGGCGCTGGACACAGGTGGCTACTTCGACCACACACCTGGGAGCGGCACGACTTC 18464
Qy 184 MetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisPro 203
Db 18463 CGTCGACGCGCATCAACGCTCGGAACAGATGGGTATTCGGTAGAGTTCAGCCACAC 18404
Qy 204 GluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAla 223
Db 18403 GAGCCGCTCTGGACAGCAGAAATGACCTTCGTTACGCCGATGGCTGACGATGGCC 18344
Qy 224 AspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetVal 243
Db 18343 GACAACATCATGACTTCGCGGTAGTCTATTCGGAGATCGCTTCCTCAAGGGATCAAG 18284
Qy 244 AlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMet 263
Db 18283 GCCACCTTCATCGCTTAAGCCTTTTACCGATCACCGCGGTGAGTATGCTTCTCAGCTC 18224
Qy 264 SerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlyTyrGly 283
Db 18223 TCCCTCTTTGAGGGGGACAC-----AATGCCCTTCTATGACGCGGTGACGAGGCTCGC 18170
Qy 284 MetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303
Db 18169 ATGTGCGCGGTGGCAAGTATTCGTCCGCGTCTTTTGACCATGCTCCGAGATCATC 18110
Qy 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
Db 18109 GCCGTGACGAACACAGTGGTTAACTCTACCGTGGCTTTCCGGTGGGGAGAGCGCTC 18050
Qy 324 ThrLysCysCysTsrPheAlaLeuAsnArgThrSerAlaPheArgLeu-----Cys 340
Db 18049 AACTACATTTGCTGGGTGCGCAATAATCGTTGACACATGTTGCTATTCGATGTATAAG 17990
Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
Db 17989 CCCGACAGGCTTCTCGGTCGGTGAGATGTTGCTTATCGATTCGTGCGCAACCCA 17930
Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
Db 17929 TATCTGGCGTATTCCTGGTACTTGTCTGCTGTTTGGACGCGATCGAGAAGGAGTACCG 17870
Qy 381 LeuProProAlaSerGlyAsnValTyr-----AsnAspLysGlu----- 394
Db 17869 CTTCCAGAGGAGGATCCGACGAGTGTGGCAGTGTTCAGCTCCGCGAGCTGACGCACTG 17810
Qy 395 ---LeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLys 413
Db 17809 GGGATCAAGCAGCTGCCAAGTCTGGGTGCTCGGATCCGCTCGATGGAGGAGTCCGAG 17750
Qy 414 MetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnVal 433

Db 17749 CTGCTGCTGAGACCTTCGCGCAACACGCTACGATTTATTCTTACCAATAAGAGGCG 17690
Qy 434 GluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
Db 17689 GAATTTGAGGATACACCGCGCAGGTGAGCCAGTTGAGCTTGCCTGATCGC 17642
RESULT 6
ABQ82091
ID ABQ82091 standard; DNA; 5500 BP.
XX ABQ82091;
XX ABQ82091;
XX 22-NOV-2002 (first entry)
XX Brevibacterium lactofermentum glnA2 and glnE encoding DNA SEQ ID NO:1.
XX Brevibacterium lactofermentum; glnA2; glnE; L-glutamine; fermentation;
XX coryneform bacterium; glutamine synthetase adenyl transferase;
XX glutamine synthetase; liver function promoting agent; enzyme; seasoning;
XX gene; ds.
XX Brevibacterium lactofermentum.
XX
FH Key Location/Qualifiers
CDS 659..1999
FT /*tag= a
FT /product= "glnA2"
FT CDS 2066..5203
FT /*tag= b
FT /product= "glnE"
XX
PN EPI1229121-A2.
XX
XX 07-AUG-2002.
XX
XX 05-FEB-2002; 2002EP-0001993.
XX
XX 05-FEB-2001; 2001JP-0028163.
XX 30-MAY-2001; 2001JP-0162806.
XX
XX (AJIN) AJINOMOTO CO INC.
XX
XX Nakamura J, Izui H, Moriguchi K, Kawashima H, Nakamatsu T;
XX Kurahashi O;
XX
XX WPI; 2002-629685/68.
XX P-PSDB; ABP53500, ABP53501.
XX
XX Coryneform bacterium which has L-glutamine producing ability and has
XX been modified so that its intracellular glutamine synthetase activity
XX should be enhanced, useful for producing L-glutamine
XX
XX Claim 11; Page 19-26; 39pp; English.
XX
XX The present invention describes a coryneform bacterium (I) which has
XX L-glutamine producing ability and has been modified so that its
XX intracellular glutamine synthetase activity should be enhanced. Also
XX described is a DNA (II) coding for a protein having glutamine synthetase
XX activity or glutamine synthetase adenyl transferase activity (see
XX /ABP53500 and /ABP53501 respectively). (I) is useful for producing
XX L-glutamine, by culturing a bacterium in a medium to produce and
XX accumulate L-glutamine in the medium and collecting the L-glutamine.
XX L-glutamine produced by (I) is useful industrially as an ingredient of
XX seasonings, as liver function promoting agents, in amino acid
XX transfections, and in comprehensive amino acid preparation. (II) is
XX useful for breeding (I). The by-product of L-glutamic acid is
XX suppressed and the production efficiency of L-glutamine is improved
XX using (II). The present sequence encodes Brevibacterium lactofermentum
XX glnA2 and glnE, which have glutamine synthetase activity and glutamine
XX synthetase adenyl transferase activity respectively.
XX
XX Sequence 5500 BP; 1241 A; 1291 C; 1590 G; 1378 T; 0 other;

xx The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium, and identifying a homologue of a gene derived
 CC from corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 xx

Sequence 1338 BP; 336 A; 367 C; 343 G; 302 T; 0 other;

Alignment Scores:

Pred. No.: 9,41e-41 Length: 1338
 Score: 470.50 Matches: 134
 Percent Similarity: 47.08% Conservative: 84
 Best Local Similarity: 28.94% Mismatches: 174
 Query Match: 19.80% Indels: 71
 DB: 22 Gaps: 12

US-10-098-602A-2 (1-454) x AAH67407 (1-1338)

Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34
 Db 85 GCCACTTGAAGTCAGTGGTGTGGCTCTCCAGAACTAGAGTCTGGTGTG----- 135
 Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysIleSerIleSerThrPhe 54
 Db 136 -----GAAGAAGGTCGATTCGAT----- 156
 Qy 55 ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp 74
 Db 157 -----GGCTCAGCCATTGAGGGGTACGCG-----CGTATCTCGAAGACGGAC 198
 Qy 75 ThrGlyTyrSerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrp 94
 Db 199 ACC-----ATTGCCGCCAGATCATCGACATTCACAGTCTCCCACTA 243
 Qy 95 SerGluGly-----AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110
 Db 244 GAAGCGGCATCTCAAACTCAGCAGCAGCCGCTGTTTCGGATGTCACGATGCCAGAC 303
 Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
 Db 304 GGACAGCCATCTTTTCTGACCCGCGCAAGTGTGCGCAGCAGGTCCTCAACTAGCTGCA 363
 Qy 131 AsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150
 Db 364 GATGAAGCTTGACCTGCATCATCTCACCAGATGATGATTTCTTTGTTG----- 414
 Qy 151 TyrLysSerAlaSerGlnLysHisIleTrpLysAsnLeuLysThr----- 164
 Db 415 -----CAAGCCTTCGACCAACGAGTCCACCTGTG 447
 Qy 165 AlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
 Db 448 CCCACTGACACCGCGGATATTCGACCAAGCCATTCATGAGGCGCGCAATTCGGT 507
 Qy 185 ArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
 Db 508 CGAAACGGATGGTAGCGCTCGAGAACTCGGCATCTCGAGTCTTCCCAACATGAA 567
 Qy 205 PheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAsp 224
 Db 568 ACTGCATGCGCAGCAAGAAATCGATTAGCCATGCGGATCGCTCACCATGCGCGAC 627
 Qy 225 ArgHisIleLeuAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValala 244
 Db 225 ArgHisIleLeuAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValala 244

Db 628 AACATCATGACCTTCGCTCATCATGAAACAGGTGGCAAGGACCAAGGCGTGGGGCA 687
 Qy 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSer 264
 Db 688 TCATTATGCCAAGCCATTCCAAAGACATGCAAGCTCCGCCATGCAACGCACATGCTCC 747
 Qy 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet 284
 Db 748 TTATTTGAGGGCGATACC-----ACGCGTTCACGATCCAGACGATCTTACATGCTG 801
 Qy 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyr 304
 Db 802 TCCAAACCCAAACAGATTCTATCGCTGGAATCTTGCATCAGCTCCAGAAATCCACGCT 861
 Qy 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324
 Db 862 GTGACCAACAGTGGGTCAATTCCTACAAAGCATCGTGTACGAAACGAGCTCCAACT 921
 Qy 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsn----- 341
 Db 922 CGGCAACCTGGGTGTATCTAATCGTTCCTGCGCTGGTTCGTGTTCTCTACCTACCGTTG 981
 Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyr 361
 Db 982 AATAAGGAGGAGTGGCGCGGTGGAGGTGCTTCTCTCATACCGCTTGTAAACCCATAT 1041
 Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGlyLysLeuGluLeu 381
 Db 1042 TTGGCGTTTTCAGTGATGCTCGGCGCTGGTGTAAAGGCATTAAGAAGGTTATGAGCTC 1101
 Qy 382 ProProProlaSerGlyAsnVal----- 389
 Db 1102 GACGAGCAGCTGAGGACGATATCTCAACTTGAGCTTCGGAACGTCGCGCCATGGC 1161
 Qy 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409
 Db 1162 TACAACGAT-----CTGCCAAGCAGCCTTGCATCAGGCACTGCGCAAAATG 1206
 Qy 410 LysGluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuIleLeuHisTyrValAsn 429
 Db 1207 GAAAGTCAGAGCTGTGTGTCATCTCGTGAGCAGCTTTTGTAGTTCCTTTCGCGC 1266
 Qy 430 AlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
 Db 1267 AATAAGTGGCGTGAATGGCGTGACTCAAGAGCAGATCACTCGTGGGAGCTCCGAAAC 1326
 Qy 450 GlyPheAsn 452
 Db 1327 AATCTTGAT 1335
 RESULT 8
 AAF71800
 ID AAF71800 standard; DNA; 1461 BP.
 XX AAF71800;
 AC AAF71800;
 XX 30-APR-2001 (first entry)
 DT
 XX
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:95.
 XX
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN W0200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00923.

```

XX 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032136.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032822.
PR 14-JUL-1999; 99DE-1032826.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX (BADI ) BASF AG.
XX Pompejus M, Kroege B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-137957/14.
XX P-PSDB; AAB79681.
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX Claim 3; Page 297-299; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX Sequence 1461 BP; 361 A; 396 C; 371 G; 333 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.08e-40 Length: 1461
Score: 470.50 Matches: 134
Percent Similarity: 47.08% Conservative: 84
Best Local Similarity: 28.94% Mismatches: 174
Query Match: 19.80% Indels: 71
DB: 22 Gaps: 12
US-10-098-602A-2 (1-454) x AAF71800 (1-1461)
QY 15 GlyLysIleAsePThrValValLeuAlaCysValAsePMetGlnGlyArgLeuMetGlyLys 34
Db 185 GGCACCTTGAAGTCAGTGGTGTGGCTCTCGAGAACTAGAGTCTCGGTG----- 235
QY 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAsePMetGlnGlyLysIleSerIleSerThrPhe 54
Db 236 -----GAAGAAGGCATCGGATTCGAT----- 256
QY 55 ValTyAlaValThrIleGluGlyIleAlaGlyGlyGlyTyArgLysSerSerValAseP 74
Db 257 -----GGCTACGCCATTGAGGGCTACGCG-----CGTATCTCGAAGCGGAC 298
QY 75 ThrGlyTyrSerAsePcysHisLeuCysAlaAsePLeuAsePLeuAsePLeuAsePLeuProTrp 94
Db 299 ACC-----ATTGCCCGCCAGATCCATCGACATTCAGGTCCTCCACATA 343
QY 95 SerGluGly---AlaValLeuAlaIleSerAsePProHisAsePValThr----- 110
Db 344 GAAGCGGGCATCTCAAAACTGCAGCAGCAGCCCTGTTTGGCATGTACGATGCCAGAC 403
QY 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
Db 404 GGACAGCCATCTTTTCTGACCGCGCAAGTGCTCGCAGCAGCAGGTCCCACTAGTGCA 463
QY 131 AseLeuLysLeuLysGlyLeuPheAlaSerGluLeuGluPheAsePLeuAsePLeuThr 150
Db 464 GATGAAGGCTTGACCTGCATGATCTCACAGAGATGAGTTCTATTGTTG----- 514
QY 151 TyrLysSerAlaSerGlnLysHisTrpLysAsePLeuLysThr----- 164
Db 515 -----CAAACCTTCGCCAACCGGATGCCACCTGTG 547
QY 165 AlaGlnProHisHisGlnTrpMetAsePLeuAsePLeuAsePLeuGluThrPheMet 184
Db 548 CCCACTGACACCGCGGATATTTCGACCAAGCCACATTCATGAGCGCGCAATTCCTG 607
QY 185 ArgSerValArgAsePLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204
Db 608 CGAAACGCGATGGTAGCGCTGGAGGAACCTCGGCATCCCTGTGCGAGTTCTCCACCATGAA 667
QY 205 PheLeuProSerGlnHisGluLeuAsePLeuAsePLeuAsePLeuMetAlaAseP 224
Db 668 ACTGCACCTGGCCAGCAAGAAATCATTTACGCCATGGGATGCCCTACCATGCCGAC 727
QY 225 ArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetValAla 244
Db 728 AACATCATGACCTTCGGCTACATCATGAAACAGGTGGCAAGGAGCAACAGCGTCGGGCA 787
QY 245 ThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsePLeuAsePLeuHisMetSer 264
Db 788 TCATTTATGCCCAAGCCATTCCAAAGACATGCGAGCTCGGCATGCACACGACATGTCC 847

```

Qy 265 LeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet 284
Db 848 TTTATTTAGGGCGATACC-----AACGGTTTCCACGATCCAGACGATCTTACATGCTG 901
Qy 285 SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyr 304
Db 902 TCCAAACCGCAACACGATTCATCGTGAATCTTGCATCAGCTCCAGAAATTCACCGCT 961
Qy 305 PhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThr 324
Db 962 GTGACCAACAGTGGTCAATCTCTACAAACGCATCGTGTACGGAACGAGCTCCAACT 1021
Qy 325 LysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuLysAsn----- 341
Db 1022 GCGCAACCTGGGGTGTATCAATCGTTCTCGCGTGGTTCGTCTACCTACCGTTTG 1081
Qy 342 SerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProThr 361
Db 1082 AATAAGAGGAGTCCGCGGGTGGAGTGGCTCTCTGATACCGCTTGTAAACCATAT 1141
Qy 362 LeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeu 381
Db 1142 TTGCGTTTTCAGTGATCTCGCGCTCGCTGTTTGAAGGCATTAAGAGGTTATGAGCTC 1201
Qy 382 ProProAlaSerGlyAsnVal----- 389
Db 1202 GACGAGCCAGCTGAGGACGATATCTCAACTTGAGCTTCCGGGAACGTCGCGCATGGC 1261
Qy 390 TyrAsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeu 409
Db 1262 TACAACCAT-----CTGCCAAGCAGCCTTGTATCAGGCACTGCGCAATG 1306
Qy 410 LysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrValAsn 429
Db 1307 GAAAGTCAGAGCTGTGTGTCGACATCTCGTGAGCAGCTTTTGTAGTTTCTTGGC 1366
Qy 430 AlaAlaAsnValGluLeuAsnGluPheSerLysGlnValThrAspTrpGluLeuAsnGln 449
Db 1367 AATAAGTGGCGTGAATGGCGTGACTACCAAGAGCAGATCACTCGTGGGAGCTCCGAAAC 1426
Qy 450 GlyPheAsn 452
Db 1427 AATCTTGAT 1435
RESULT 9
AAH68531/c
ID AAH68531 standard; DNA; 349980 BP.
XX
AC AAH68531;
XX
DT 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 7066.
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EF1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;
Alignment Scores:
Pred. No.: 4,55e-37 Length: 349980
Score: 470.50 Matches: 134
Percent Similarity: 47.08% Conservative: 84
Best Local Similarity: 28.94% Mismatches: 174
Query Match: 19.80% Indels: 71
DB: 22 Gaps: 12
US-10-098-602A-2 (1-454) x AAH68531 (1-349980)
Qy 15 GlyLysIleAspThrValValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLys 34
Db 264071 GGCACCTTGAAGTCAGTGGTGTGGCTCTCTGCAGAACATAGAGTCGCGTTG----- 264021
Qy 35 ArgLeuThrGlyArgHisPheLeuGlyLeuAspGlnLysLysIleSerIleSerThrPhe 54
Db 264020 -----GAAGAAGCATCGGATTCGAT----- 264000
Qy 55 ValTyrAlaValThrIleGluGlyIleAlaGlyGlyTyrGluIleSerSerValAsp 74
Db 263999 -----GGCTCAGCCATTGAGGGCTACGCG-----CGTATCTCGAAGCGGAC 263958
Qy 75 ThrGlyTyrSerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrp 94
Db 263957 ACC-----ATTGCCCGCCAGATCCATCCAGATCCAGGTCCTCCCACTA 263913
Qy 95 SerGluGly---AlaValLeuAlaIleSerAsnProHisAsnPheValThr----- 110
Db 263912 GAAGCGGGCATCTCAAAACTGCAGGAGCAGCGCTGTGTTGCGATGTCCAGATGCCAGAC 263853
Qy 111 SerGluProLeuPheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAla 130
Db 263852 GGACAGCATCTTTTCTGACCCCGCCCAAGTCTGCGCAGGCGGTCCTCAACTAGCTGCA 263793
Qy 131 AsnLeuLysLysGlyLeuPheAlaSerGluLeuGluPheAsnLeuPheAsnGluThr 150
Db 263792 GATGAAGGCTTGACCTGCATGATCTCACCAGAGATGTGATTTCTATTGGTG----- 263742
Qy 151 TyrLysSerAlaSerGlnLysHisTrpLysAsnLeuLysThr----- 164
Db 263741 -----CAAAGCCTTCGCAACCAACGAGCATGCCACCTGTG 263709
Qy 165 AlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMet 184
Db 263708 CCCACTGACACCGCGGATATTTTCACCAAGCAGCATTTCAATGAGGCGCGCAATTCGT 263649
Qy 185 ArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGlu 204


```
QY 81 HisLeuCysAlaAspLeuAenSerLeuHisLeuLeuProTyrSerGluGly----- 97
Db 105597 ATGTTCAACCCGATCCCTCGAGATCTTCAGATCTTCGGTGGCGGTGGCCCGCAAGT 105656
QY 98 -----AlaValLeuAlaIleSerAenProHisAenPheValThrSerGluProLeu 114
Db 105657 ACCGCCCGCATGTTCTGCACATCTGACCCCG-----GACGGCGAGCCTCG 105704
QY 115 PheCysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAenLeuLysLeu 134
Db 105705 CTCGGCGATCCACGTCACGTTGTAAGCGC-----ACGTCCGCAAGCAAGAGAC 105755
QY 135 LysGlyLeu-----PheAlaSerGluLeuGluPheAenLeuPheAenGluThrTyr 151
Db 105756 AAGGGCTTCAGTTCAGTTCGTCGTAATCGAGTTCTACCTGTTCGAA----- 105806
QY 152 LysSerAlaSerGlnLysHisIleThrAenLeuLysThrAlaGlnProHisHisGlnTyr 171
Db 105807 -----AGCCAGGACGACTGCTCAAGGCCCGCCGATCGACGAGGCGGCTAC 105857
QY 172 Met---AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAenLys 190
Db 105858 TTGACACACGTGCGCGCACCCAGCGATGGAC---TTCGCCGTGCCACCGTGAACATG 105914
QY 191 LeuGluGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
Db 105915 CTCGACGAATGGCATCTCGGTGGATATCTCGCACCATGAGCGGGCCAGCCAGAAC 105974
QY 211 GluLeuAenPheValProAlaAspProLeuThrMetAlaAspArgHisIleLeuAlaLys 230
Db 105975 GAGATCGATCTGCGATACGCGGACGCGCTCACCATGCGCGCACCAACATCATGATCCGT 106034
QY 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
Db 106035 ACGGTGTCAGGAGATCTCTTGGAGCGTGGATTCACGCCAGCTCATGCCCAAGCCC 106094
QY 251 SerSerThrAlaLeuGlyAenAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
Db 106095 CTGCGCCAGCGCGCGTTCGGCATGCGACACGCACTGAGCCTG-----TTCGAGGC 106148
QY 271 GluLysAenAlaPheTyrAspGlnAenAspGluTyrGlyMetSerThrLeuAlaAenAsn 290
Db 106149 GATTCCAACGCCCTTACGAGCGCGGTTCAGGAGTTCAACTGTCCTGACCGCGCCGAC 106208
QY 291 TrpIleAlaGlyLeuLysTyrValProGluAlaThrTyrPheAlaSerTyrIle 310
Db 106209 TTCGCCCGCGCATCTGTACCATCGCGCGCAATCTGCGCGTCAACGACAGTACGTC 106268
QY 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysCysTrpAlaIle 330
Db 106269 AACTGTACAGCGCCTGTGGCGGTGTAACGAGCGCACCAAGCTACATCTGCTGGGCGCAC 106328
QY 331 AspAsnArgThrSerAlaPheArgLeuCysAen-----SerLysSerGluGlyIle 347
Db 106329 AACAAACGTTCCGCGCTGCTGCTGATCCCGCAGTACAAAGCGGCAAGCACTCCGCG 106388
QY 348 AsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIle 367
Db 106389 CGATGAGTTCGCTGCGCTGACCCCGCGTGGCGCAACCCATACCTCTACTCGGTGCTG 106448
QY 368 IleAlaAlaGlyIleSerGlyIleGluLysLeuGluLeuProProAlaSerGly 387
Db 106449 CTGCGCGCGCTCTCGAGGTATCGATGATGACAGCATGCTGGCGGCGCTACCATGTAT 106508
QY 388 AsnValTyr-----AsnAspLysGlu-----LeuProGluPheProAsn 400
Db 106509 GACGTGTGGAGCTCACCGACGCGCAAGCTGAGCTATGGCATTCAGCGCTGCGCGCGC 106568
QY 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGlySerLysMetLeuAsnLysThrPheGly 420
Db 106569 TCCTGGACGAGCGCTGAAGATCATGGAGAAATCCGACTTCGTGTCGCGCGACGTGCTCGGC 106628
QY 421 GluLysLeuIleLeuHisTyrValAsnAlaAlaAenValGluIleAenGluPheSerLys 440
```

```
Db 106629 GAACACGCGTTCGGTACTTCTCTGACAAACAGCACCGAATGGAGGAGTACAACCCAG 106688
QY 441 GlnValThrAspTrpGluLeuAsnGln 449
Db 106689 CAGGTACGCGCTTACGAGCTGAAGAAG 106715

RESULT 12
AAV04201
ID AAV04201 standard; DNA; 4259 BP.
XX
AC AAV04201;
XX
DT 14-APR-1998 (first entry)
XX
DE Merged contigs pFI913/PCR product/pFI901.
XX
KW Merged contigs; pFI913/PCR product/pFI901; vanillin production;
KW vanilla pod; vanillic acid production; food flavouring; ss.
XX
OS Pseudomonas fluorescens.
XX
PN WO9735999-A2.
XX
PD 02-OCT-1997.
XX
PF 24-MAR-1997; 97WO-GB00809.
XX
PR 23-MAR-1996; 96GB-0006187.
XX
PA (FOOD-) FOOD RES INST.
XX
PI Gasson MJ, Narbad A, Rhodes MJ, Walton NJ;
XX
DR WPI; 1997-489659/45.
XX
PT Production of vanillin from trans-ferulic acid using 3 enzymes -
PT optionally followed by oxidation to vanillic acid, useful as a food
PT flavouring
XX
PS Example 5; Fig 19; 127pp; English.
XX
CC The present sequence was used in the development of a novel method
CC for the production of vanillin.
CC The method comprises exposing trans-ferulic acid, or its salts, to
CC trans-ferulate:CoASH ligase, trans-feruloyl SCoA hydratase and
CC 4-hydroxy-3-methoxyphenyl-beta-hydroxypropionyl-S-CoA cleavage
CC activities. A similar novel method in which vanillin is converted
CC to vanillic acid, or its salts, utilises a 4th enzyme, especially
CC vanillin dehydrogenase. Vanillin is a well known food flavouring,
CC and vanillic acid can be used to make oligomers, or as a monomer
CC for production of polyesters. By expressing vanillin in plants or
CC food-grade microorganisms, vanillin is generated in situ during
CC preparation of fermented foods and beverages. Transgenic plants,
CC particularly food plants, may be able to produce other
CC flavour/aroma compounds in addition to vanillin. This method is
CC less expensive than extraction from vanilla pods, and generates a
CC natural product as only enzymes are used. The product is very
CC similar to natural Vanilla pod extract.
XX
SQ Sequence 4259 BP; 965 A; 1297 C; 1215 G; 782 T; 0 other;

Alignment Scores:
Pred. No.: 1,188-37 Length: 4259
Score: 449.00 Matches: 128
Percent Similarity: 46.05% Conservative: 82
Best Local Similarity: 28.07% Mismatches: 224
Query Match: 18.90% Indels: 22
DB: 18 Gaps: 6

US-10-098-602A-2 (1-454) x AAV04201 (1-4259)
QY 1 MetThrIleThrTyrAspGluLeuAenAenLeuIleArgAenGlyLysIleAspThrVal 20
```


417 ATGAGTAAACACCTCGACGACCTCAGGATTGGTTGAAGACCAACAGATCACAGAAGTC 476
 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
 477 GAATGTCATGATGGCGACTTACCGGATCATCCCGCGCAAGATCTCGCCCAACCAACAG 536
 41 PheLeuGlyLeuAspGlnLysLysLeuSerThrPheValThrAlaValThrIle 60
 537 TTCATTGCGGAAAAAGGATCGCGCTGCCGAGAGTGTGTTGTCAGACAGTGCACGGC 596
 61 GluGlyIleAlaGlyGlyGlyThrGluIleSerValAspThrGlyTyrSerAspCys 80
 597 GACTATGTGCAAGACGACATCTAT---TACGAATGCTCGACCGCGCGACATCGACATG 653
 81 HisLeuCysAlaAspLeuAsnSerLeuHisLeuProTyrPheSerGluGlyAlaValLeu 100
 654 ATTCGCGCCCGACCAACGCGGTGTTCTCTGTCATCGGCATCGGCATCGACCGACCGC 713
 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
 714 CAGGTGATTACGACACCTACGACCAAGCAGGCAACCGATCGAGCTGTGCGCCACGCAAC 773
 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSer 140
 774 GTCCTCAAGAAAGTCTCAAACTCTATTCCGACAAAGGCTGCGCAGCGATCGTGCGCGC 833
 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisThrLys 160
 834 GAATGAGTCTACCTG-----ACCAAGCGATGACGACCGGATTAC----- 878
 161 AsnLeuLysThrAlaGlnPro-----HisHisGln 170
 879 -----CCATTGCAACGCGCGTGGCGGTTCCGACGTCGCGAAATCGGTGCGCAA 929
 171 TrpMetAsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLys 190
 930 TCGTTCTCTATCGAAGCGGCAACGAATTCACCGCGTGTTCGAAGACGCTACGACTGG 989
 191 LeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHis 210
 990 TCGAAGTCCGAGGAGCTGATCTCGATCGCTGATCCACGAGACGCGCGCGCAGATG 1049
 211 GluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleAlaLys 230
 1050 GAAATCAACTTCGCTCAGCGCGACGCGCTGTCCTCGCGCACGACGATCTCGTGTTCAG 1109
 231 HisGlyValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeu 250
 1110 CGCACCATGCGGAGCGCGCTCAAGCACAACGTCGCGCGCATCTCATGTGCCAAGCGC 1169
 251 SerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThr 270
 1170 ATGACCGCGGAGCTCGCGCGCATGCTGACCTGACGAGCATCATCGATATCGAGACC 1229
 271 GluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMetSerThrLeuAlaArgAsn 290
 1230 GGCAAGAAGCGTTC-----TCCAATGAAGCGGAGCATGACCGAGTTGTTCTCTAAC 1283
 291 TrpIleAlaGlyLeuLeuLysTyrValProGluAlaThrTyrPhePheAlaSerTyrIle 310
 1284 CACATCGCGCGCTCGCAAAATTCATCCCTGAATGCTGCGCGTGTTCGCGCGCCCAACGTC 1343
 311 AsnSerTyrLysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrTrpAlaIle 330
 1344 AACTGTTGCGCGCTTCCTGCGCGACATTCGCGCGCGGTGAACGTCGAGTGGCGGCA 1403
 331 AspAsnArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGlu 350
 1404 GAAACCGTACCGTGGCGCTCGGGTCCGGATCGCGGCTCTCAAAACCGTGGGTGGA 1463
 351 LeuArgIleGlyAlaAspLeuAsnProTyrLeuAlaPheSerAlaIleAlaAla 370

1464 AACCGCTCGCGGTGCGGACGCAACCCCTACTCTGGCGATTGCCGAGCGCTGCTGTC 1523
 371 GlyIleSerGlyIleGluGlyLysLeuGluLeuProProAlaSerGlyValAsnValTyr 390
 1524 GGCTACATCGCATGCTCGAAGGTATCAACCAAGCGCGCTGTGGTGGGTGCTGTGTTAC 1583
 391 AsnAspLysGluLeuProGluPheProAsnSerLeuGlnAsnAlaThrHisLeuLys 410
 1584 GAGCGCGCAACCTG---CGTCTGCGCTGACCATCGAAGCGCTCTGGAACGATCGAA 1640
 411 GluSerLysMetLeuAsnLysThrPheGlyGlyLysLeuLysLeuHisTyrValAsnAla 430
 1641 AACGCAAGACCATCGAGAAATACCTGGGTCAACAATTCATCTGCTAGTCGCGGTC 1700
 431 AlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGlu 446
 1701 AAGCGGCGGACGATGAAATCTTCAAGCGGCTGATCAGTCATGGA 1748
 RESULT 13
 ABN68493
 ID ABN68493 standard; DNA; 1344 BP.
 XX AC ABN68493;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 4899.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus agalactiae.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB04789.
 XX PR 27-OCT-2000; 2000GB-0026333.
 XX PR 24-NOV-2000; 2000GB-0028727.
 XX PR 07-MAR-2001; 2001GB-0005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Ros VI, Grandi G, Fraser C;
 XX PI Tettelin H;
 XX DR WPI; 2002-352536/38.
 XX DR P-PSDB; ABP27862.
 XX PT New Streptococcus protein for the treatment or prevention of infection
 XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
 XX PT for detecting a compound that binds to the protein -
 XX PS Claim 7; Page 3651-3652; 4525pp; English.
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX Sequence 1344 BP; 426 A; 224 C; 277 G; 417 T; 0 other;

Alignment Scores:

Pred. No.: 5,51e-38 Length: 1344
 Score: 445.00 Matches: 131
 Percent Similarity: 44.14% Conservative: 80
 Best Local Similarity: 27.41% Mismatches: 213
 Query Match: 18.73% Indels: 54
 DB: 24 Gaps: 10

US-10-098-602A-2 (1-454) x ABN68493 (1-1344)

QY 1 MetThrIleThrTyrAspGluLeuAsnAsnLeuIleArgAsnGlyIleAspThrVal 20
 DB 1 ATGACTATCACAGCAGAGATATTCGCGAAGTTAAGGAAAAAATGTTACCTTCCTC 60
 QY 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyIleAspThrGlyArgHis 40
 DB 61 CGTTTGATGTTACAGATATCTAGGAGTAATGAAAAATGTCGAAATTCG----- 111
 QY 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla 57
 DB 112 -----GCTACAGATGACAAATAGACAAAGTTTATCAATAAAGCAATGTTGATGGA 165
 QY 58 ValThrIleGluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyr 77
 DB 166 TCTTCTATTGAGGCTTT-----GTTGCTATCAATGAA 198
 QY 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGlu--- 96
 DB 199 TCGATATGATTTATATCTGATTTAGTACTGTTGATGTTCTTCTTGGGAGATGAA 258
 QY 97 ---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115
 DB 259 AATGGAGCTGTTCCAGGCTCTAATCTGTGATATCTATACAGCAGAGAGAACCTTTTGGC 318
 QY 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLys 135
 DB 319 GGAGATCCACGTCGGAATCTTAAGCGTATATGAAACGTTATGCAAGAGATGGGCTATAAG 378
 QY 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAla 154
 DB 379 TCATTTAACTAGGACAGAACCTGAATTTTCTCTATT----- 417
 QY 155 SerGlnLysHisTrpLysAsnLeuLysThrAlaGlnProHisGlnTrpMetAsnIle 174
 DB 418 -----AAATGGATGAAATGGTAACCAACCAACACCTTGATGTT 453
 QY 175 SerAlaSerSerGly-----IleGluThrPheMetArg 185
 DB 454 AATGACAAAAGGAGCTACTTTGACTTAGCTCCACAGACTTAGCAGATATAACAGTCGT 513
 QY 186 SerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisProGluPhe 205
 DB 514 GAATTTGTAATGCTTTACACAAATGGGCTTTGAAGTTGAAGCTAGTCATCATGAAGTA 573
 QY 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225
 DB 574 GCAGTTGGACAGCATGAAATGACTTTAAATATGATGATGTTTAAAGCCTGTGACAAAC 633
 QY 226 HisIleAlaLysHisGlyValArgGluMetAlaGluSerGlyMetValIleThr 245
 DB 634 ATTCAATTTTAAATGGTTGTTAAACAAATTCGTCGTAAGCATGGTTTATATGCTACT 693
 QY 246 PheMetAlaLysLeuSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
 DB 694 TTTATGCGAAAACCTAAGTTTGGTATCAATGGTTAGGTATGCAATGTTGATATGCTCT 753
 QY 266 GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGlyMet--- 284

DB 754 TTTGATAATGAA---GGGAACAATGCTTTCTTTGATCCAGAAGATCCACGTGGAATGCAA 810
 QY 285 ---SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303
 DB 811 TTATCAGAGGATGCTACTACTTTTAGGTGGTTTAAATGAACACGCTTACAATTACACA 870
 QY 304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
 DB 871 GCTATCATTAATCAACTGTTAAATCTTATAACGTTTGGTACCAGGATACGAACACCA 930
 QY 324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
 DB 931 GTTTACGTTGCTGGGAGGTCGCAATCGTTCGCCACTTATTCTGTGTACCTGCATCTCGA 990
 QY 344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363
 DB 991 GGTATGGGAACACGCTAGAAATTCGCTCGGTTGACCCACAGCAAAATCCTTATCTTGT 1050
 QY 364 PheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGlyLysLeuGluLeuProPro 383
 DB 1051 TTATCAGTACTTTTGGGATCAGGACTTCAGGGAATCGAATAAATTTGACACACAGNA 1110
 QY 384 ProAlaSerGlyAsnValTyr-----AsnAspLysGluLeuPro 396
 DB 1111 CCAATTGAGACAAATATCTATGACTGTAGAAGAACGTCGTCGCAAGCAGGTATTGTT 1170
 QY 397 GluPheProAsnSerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsn 416
 DB 1171 GATTTACCATCTACTCTTCATAATGCTCTTGAAGCAGCTTGAAGAAGATGAAGTGGTAAA 1230
 QY 417 LysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsn 436
 DB 1231 GCGGCCCTAGGAACCTCATATCTATCTAATTTCTTAGACGCTAAGCGTATTGAATGGCA 1290
 QY 437 GluPheSerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
 DB 1291 AGTTATGCGACTTATGTTTCAATGGGAATTTGATAATTACCTAGATTATAT 1344

RESULT 14

ABN71527/c
 ID ABN71527 standard; DNA; 2155561 BP.

XX AC ABN71527;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 10967.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX OS Streptococcus sp.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX XX

PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -

PS Claim 8; Page 4196-4488; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Alignment Scores:

Pred. No.: 4.23e-33 Length: 2155561
 Score: 445.00 Matches: 131
 Percent Similarity: 44.14% Conservative: 80
 Best Local Similarity: 27.41% Mismatches: 213
 Query Match: 18.73% Indels: 54
 DB: 24 Gaps: 10

US-10-098-602A-2 (1-454) x ABN71527 (1-2155561)

Qy 1 MetThrIleThrTyAspGluLeuAsnLeuLeuArgAsnGlyLysIleAspThrVal 20
 Db 1726056 ATGACTATACAGCAGAGATATCGTCGCGAAGTAAAGAAAAATGTTACCTCCTC 1725997
 Qy 21 ValLeuAlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHis 40
 Db 1725996 CGTTTGATGTCACAGATATCTTAGGAGTAAATGAAAAATGCGAAATTCGCG- 1725946
 Qy 41 PheLeuGlyLeuAspGln-----LysLysIleSerIleSerThrPheValTyrAla 57
 Db 1725945 -----GCTACAGATGACAAATTAGACAAAGTTTATCAATAAAGCAATGTTGATGGA 1725892
 Qy 58 ValThrIleGluGlyIleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyr 77
 Db 1725891 TCTTCTATTGAGGCTTT-----GTTGCTATCAATGAA 1725859
 Qy 78 SerAspCysHisLeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGlu--- 96
 Db 1725858 TCTGATATGATTTATATCTCTGATTTAGATATCTTGGATGTTGCTTCTTGGGAGATGAA 1725799
 Qy 97 ---GlyAlaValLeuAlaIleSerAsnProHisAsnPheValThrSerGluProLeuPhe 115
 Db 1725798 AATGGAGCTGTTGCGAGTCTAATCTGTGATATCTATACAGCAGAGAGAACCTTTTGCC 1725739
 Qy 116 CysSerProArgValIleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuLysLys 135
 Db 1725738 GGAGATCCACGTCGAAATCTTAAGCGTAATATGAACGATGCAAGAGATGGGCTATAAG 1725679
 Qy 136 GlyLeu---PheAlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLysSerAla 154
 Db 1725678 TCATTTAACTTAGACACGACCACTGAAATTTTCTCTATTT----- 1725640
 Qy 155 SerGlnLysHisTyrPheAsnLeuLysThrAlaGlnProHisHisGlnTrpMetAsnIle 174
 Db 1725639 -----AAATGGATGAAATGTAACCAACACTTGATGTT 1725604
 Qy 175 SerAlaSerSerGly-----IleGluThrPheMetArg 185

Db 1725603 AATGACAAAGGAGGCTACTTTGACTTCCAAACAGACTTAGCAGATAATACACGTCGT 1725544
 Qy 186 SerValArgAsnLysLeuGluAlaGlyLeuMetGluAlaThrHisProGluPhe 205
 Db 1725543 GAAATGTTTAAATGCTTAAACAAATGGCTTTGAAGTGTGAAGCTAGTCATCATGAAGTA 1725484
 Qy 206 LeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAlaAspArg 225
 Db 1725483 GCAGTTGGACAGCATGAATTTGACTTTAAATATGATGATGATGTTTAAAGCCTGTGACAA 1725424
 Qy 226 HisIleIleAlaLysHisGlyValArgGluMetAlaGlnSerGlyMetValAlaThr 245
 Db 1725423 ATTCAAATTTAAATTTGTTTAAACAAATGCTGCTGAAGCATGTTTATATGCTACT 1725364
 Qy 246 PheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeu 265
 Db 1725363 TTTATGCAAAACCTTAAGTTTGGTATCAATGGTTTCAGGTATGCTAATATATGCTCTT 1725304
 Qy 266 GlnAspAlaGluThrGluLysAsnAlaPheTyAspGlnAsnAspGluTyrGlyMet--- 284
 Db 1725303 TTTGATAATGAA---GGGAACAATGCTTCTTTGATCCAGAGATCCACGTGGAATGAA 1725247
 Qy 285 ---SerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303
 Db 1725246 TTATCAGAGATGCTACTACTTTTATAGTGGGTTAATGAACACGCTTACAATTACACA 1725187
 Qy 304 TyrPhePheAlaSerTyIleAsnSerTyLysArgGluGlnProLeuThrPheAlaPro 323
 Db 1725186 GCTATCATTTAATCCAACTGTTAATTTCTTATAAACGTTTGTGTACAGGATACGAAGCACCA 1725127
 Qy 324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
 Db 1725126 GTTTACGTTGCTTGGGACGCTCGCAATCGTTGCCACTTATTCGTGTACCTGCTCTCGA 1725067
 Qy 344 SerGluGlyIleAsnValGluLeuArgGlyGlyAlaAspLeuAsnProTyrLeuAla 363
 Db 1725066 GGTATGGAAACACGCTAGAAATTCGCTCGCTGACCCAAACAGCAAAATCCTTATCTGCT 1725007
 Qy 364 PheSerAlaIleIleAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProPro 383
 Db 1725006 TTATAGTACTTTTGGATCAGGATGAGGGAATCGAGAAATAAATTTGAAGCACCAAGAA 1724947
 Qy 384 ProAlaSerGlyAsnValTyr-----AsnAspLysGluLeuPro 396
 Db 1724946 CCAATTGAGCAAAATATCTATGCTATGCTGTAGAGAAGCTGCTCAGCAGGTATGTT 1724887
 Qy 397 GluPheProAsnSerLeuGlnAsnAlaThrHisLeuLysGlySerLysMetLeuAsn 416
 Db 1724886 GATTTACCATCTACTCTTCATAATGCTTCTTGAAGCACTTGAAGAAGATGAAGTGAATAA 1724827
 Qy 417 LysThrPheGlyGluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsn 436
 Db 1724826 GCGCCCTAGGAACCTCATATCTATATAATTTCTTAGACGCTAGCGTATGTAATGGCA 1724767
 Qy 437 GluPheSerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
 Db 1724766 AGTTATGCGACTTATGTTTCACAATGCGAAATGTAATTAATCCTAGATTATAT 1724713

RESULT 15

ABN68494
 ID ABN68494 standard; DNA; 1344 BP.

XX AC ABN68494;

XX DT 01-JUL-2002 (first entry)

XX DE Streptococcus polynucleotide SEQ ID NO 4901.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.


```

Db      1249  ATCTACACTAATTTCTTAGAAGCAAAACGAATTGAATGGTCTCTCTATGCAACTTTTGT 1308
Qv      443  ThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db      1309  TCTCAATGGGAATTGACCATTATATTCATAATTAT 1344

```

Search completed: December 18, 2003, 00:52:31
Job time : 7163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:44:55 ; Search time 31 Seconds
(without alignments)
2735.110 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLIRNGKIDTV.....INFEKSQVTDWELNQGPNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	773.5	32.6	454	15	US-10-156-761-14253	Sequence 14253, A
2	475.5	20.0	446	15	US-10-062-458-2	Sequence 2, Appl1
3	470.5	19.8	446	10	US-09-738-626-5942	Sequence 5942, Ap
4	432	18.2	448	9	US-09-815-242-13378	Sequence 13378, A
5	427.5	18.0	453	15	US-10-156-761-13481	Sequence 13481, A
6	424.5	17.9	446	9	US-09-815-242-5733	Sequence 5733, Ap
7	424.5	17.9	446	9	US-09-815-242-12338	Sequence 12338, A
8	424.5	17.9	446	9	US-09-815-242-12775	Sequence 12775, A
9	402	16.9	451	9	US-09-815-242-10514	Sequence 10514, A
10	322	13.6	477	10	US-09-738-626-6985	Sequence 6985, Ap
11	291	12.2	469	15	US-10-156-761-13532	Sequence 13532, A
12	287	12.1	472	9	US-09-815-242-11137	Sequence 11137, A
13	281.5	11.8	500	15	US-10-187-267A-57	Sequence 57, Appl
14	265.5	11.2	469	9	US-09-815-242-12088	Sequence 12088, A
15	265	11.2	481	9	US-09-815-242-11507	Sequence 11507, A

16	247	10.4	469	9	US-09-815-242-13789	Sequence 13789, A
17	241	10.1	468	12	US-10-299-799-1	Sequence 1, Appl1
18	241	10.1	469	9	US-09-815-242-10418	Sequence 10418, A
19	238.5	10.0	306	15	US-10-156-761-7750	Sequence 7750, Ap
20	106	4.5	373	15	US-10-205-823-152	Sequence 152, App
21	96	4.0	313	12	US-10-306-762-211	Sequence 211, App
22	95.5	4.0	522	12	US-10-094-749-3166	Sequence 3166, Ap
23	94.5	4.0	728	9	US-09-815-242-5216	Sequence 5216, Ap
24	94	4.0	606	10	US-09-771-161A-240	Sequence 240, App
25	94	4.0	606	10	US-09-771-161A-241	Sequence 241, App
26	94	4.0	651	9	US-09-870-937-10	Sequence 10, Appl
27	94	4.0	651	10	US-09-974-298-112	Sequence 112, App
28	94	4.0	651	12	US-10-354-358-8	Sequence 8, Appl1
29	93.5	3.9	6281	9	US-09-815-242-12996	Sequence 12996, A
30	93	3.9	1455	12	US-10-238-075-935	Sequence 935, App
31	92	3.9	390	11	US-09-214-592-28	Sequence 28, Appl
32	92	3.9	2086	9	US-09-815-242-5639	Sequence 5639, Ap
33	92	3.9	5795	9	US-09-815-242-12610	Sequence 12610, A
34	91	3.8	1050	11	US-09-954-987B-175	Sequence 175, App
35	91	3.8	1050	12	US-10-272-502A-22	Sequence 22, Appl
36	90	3.8	521	15	US-10-081-051-60	Sequence 60, Appl
37	89	3.7	2677	12	US-10-144-194A-22	Sequence 22, Appl
38	88	3.7	531	15	US-10-205-823-82	Sequence 82, Appl
39	87.5	3.7	441	12	US-10-396-122-79	Sequence 79, Appl
40	87.5	3.7	1185	9	US-09-815-242-11466	Sequence 11466, A
41	87	3.7	1294	9	US-09-815-242-13724	Sequence 13724, A
42	86.5	3.6	605	9	US-09-782-980-62	Sequence 62, Appl
43	86.5	3.6	1634	12	US-10-435-766-23	Sequence 23, Appl
44	86	3.6	850	10	US-09-738-626-5785	Sequence 5785, Ap
45	85.5	3.6	1185	10	US-09-895-913A-246	Sequence 246, App

ALIGNMENTS

RESULT 1

US-10-156-761-14253

; Sequence 14253, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156.761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 14253

; LENGTH: 454

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-14253

Query Match 32.6%; Score 773.5; DB 15; Length 454;

Best Local Similarity 38.7%; Pred. No. 7.8e-72; Indels 13; Gaps 8;
Matches 175; Conservative 81; Mismatches 183;

Qy 6 DELNLIRNGKIDTVVLACVDMQRLMGKLTGRHFLGLDQKKISIST----FVAVATIE 61

Db 11 BELHALVAGGEIDTVLAFDPDMQRLQGRKFAARFLL---DEVLEHGTGECNYLLAVD 67

Qy 62 GIAGGGVEISSVDTGYSDDLCLADNLHLLPWSGAVLAISNPHNFVTEPLFCSPRVI 121

Db 68 MNTVDGYDMSWRGDFAMHPDLSTLRQVWPNAGTALLIAD-LAWNDSPPVVAAPRQI 126

QY 122 LMOQIERLANLKLGLFASLEFNLNETYKSAQKHWNKLTAKOAPHQWMMNISASSGIE 181
Db 127 LRQLDLAALGYTAQVGTLEFIVFDYEQAWDAGYKGLTPANQYNDYSLVLTGRIE 186
QY 182 TMRSVRNKLEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIIAKHGVREMAEQSG 241
Db 187 PLLRRIRNEMQAAGLTVESAGKCNPCQHEIVPRYDEALVTCDOHAVYKTKAKEIAAQEG 246
QY 242 MWATFNKLSSTALGNACHHMSLQDAETKNAFYQNDYGNMSTLARNWIAGLLYKYPE 301
Db 247 VSTTFWAKYNRE-GNSCHILSLADAD-GTNAMAGDGPC-GMSDYMRFHLAGQLAALRD 303
QY 302 ATFFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSEGINVELRIGGADLNPY 361
Db 304 FSLLYAPNINSYKRFQGFAPTAVANGYDNRTCALRVV-GHGRSWMFENRLPGGDVNH 362
QY 362 LAFSAIIAAGISGIEKLELPPPPASGNVYNDKELPEFPNLSQNAATHLLKESKMLNKTGFE 421
Db 363 LAVAGLVAAGLYGIEHKLPEACAGNAV-AAEYEHVPTTLREAELWENSPIAKAAGD 421
QY 422 KLILHYVNAANVINEFSQVTDWELNQGNR 453
Db 422 DVVAHYRNMARVELKAFDAAVTDWELRRSFER 453

RESULT 2

US-10-062-458-2
; Sequence 2, Application US/10062458
; Publication No. US20030003550A1
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: IZUI, HIROSHI
; APPLICANT: MORIGUCHI, KAYO
; APPLICANT: KAWASHIMA, HIROKI
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMINE BY FERMENTATION AND L-GLUTAMINE
; FILE REFERENCE: 219181US0
; CURRENT APPLICATION NUMBER: US/10/062,458
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: JP 2001-28163
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: JP 2001-162806
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-10-062-458-2

Query Match 20.0%; Score 475.5; DB 15; Length 446;
Best Local Similarity 29.2%; Pred. No. 1.2e-40;
Matches 135; Conservative 83; Mismatches 174; Indels 71; Gaps 12;
QY 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISITFVYAVTIEGIAGGYSVD 74
Db 29 GHLSVVPVPAELESAL-----BEGIGFD-----GSAIEGYA-----RISEAD 66
QY 75 TGYSCHLCADLNSLHLLPWSEG-AVLAINPHNFVT---SEPLFCSPRVILMQIERIA 130
Db 67 T-----IARPDPTFQVLPLEAGISKLAARLFCVDTMPDQGFSDPRQVLRQVQLAA 121
QY 131 NLKGLFASLEFNLNETYKSAQKHWNKLT-----AOPHHQWMMNISASSGIE 184
Db 122 DEGLTCMISPEIEFYLV-----QSLRTNGLPPVPTDNGGYFQATNEAPNR 169
QY 185 RSVRNKLEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIIAKHGVREMAEQSGMVA 244
Db 170 RNAMVALEELGIPVEFSHHTAPGQOEIDL RHADALTADNIMTFRYIMKQVARDQGVGA 229

QY 245 TFMAKLSSTALGNACHHMSLQDAETKNAFYQNDYGNMSTLARNWIAGLLYKYPEATY 304
Db 230 SPMPKPFQEHAGSAMHTMSLPEGDT--NAPHPDDSYWLSKTAQFIAGILHHAPEFTA 287
QY 305 FPAVSINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCN-----SKSEGINVELRIGGADLNPY 361
Db 288 VTNOWNSYKRIYVYNEAPTAATGWGNSRSALVRVPTVRLNKESRRRVEVRLPDTACNPY 347
QY 362 LAFSAIIAAGISGIEKLELPPPPASGNV-----YNDKELPEFPNLSQNAATHLL 409
Db 348 LAFSVMLGAGLKGIEGYELDEPAEDDTSNLSFRERRANGYND-----LPNSLDQALROM 402
QY 410 KESKMLNKTGFEKLILHYVNAANVINEFSQVTDWELNQGN 452
Db 403 EKSELVADILGEHVFEFFLRNKNRWRDYOEQITPWLNRNLD 445

RESULT 3

US-09-738-626-5942
; Sequence 5942, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 5942
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5942

Query Match 19.8%; Score 470.5; DB 10; Length 446;
Best Local Similarity 28.9%; Pred. No. 4.1e-40;
Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;
QY 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISITFVYAVTIEGIAGGYSVD 74
Db 29 GHLSVVPVPAELESAL-----BEGIGFD-----GSAIEGYA-----RISEAD 66
QY 75 TGYSCHLCADLNSLHLLPWSEG-AVLAINPHNFVT---SEPLFCSPRVILMQIERIA 130
Db 67 T-----IARPDPTFQVLPLEAGISKLAARLFCVDTMPDQGFSDPRQVLRQVQLAA 121
QY 131 NLKGLFASLEFNLNETYKSAQKHWNKLT-----AOPHHQWMMNISASSGIE 184
Db 122 DEGLTCMISPEIEFYLV-----QSLRTNGLPPVPTDNGGYFQATNEAPNR 169
QY 185 RSVRNKLEAGILMEATHPEFLPSQHELNFVPADPLTMADRHIIAKHGVREMAEQSGMVA 244
Db 170 RNAMVALEELGIPVEFSHHTAPGQOEIDL RHADALTADNIMTFRYIMKQVARDQGVGA 229
QY 245 TFMAKLSSTALGNACHHMSLQDAETKNAFYQNDYGNMSTLARNWIAGLLYKYPEATY 304
Db 230 SPMPKPFQEHAGSAMHTMSLPEGDT--NAPHPDDSYWLSKTAQFIAGILHHAPEFTA 287

Qy 305 FFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCN---SKSEGINVELRIGGADLNPY 361
Db 288 VTNQWNSYKRVVYNEAPTAAATGVSNSRSLVRVPTVRLNKESRRVEVRLPDTACNPF 347
Qy 362 LAFSAITAGISGIEEKLLEPPASGNV-----YNDKELPEFPNSLQNAHLL 409
Db 348 LAFVMLGAGUGIKIGEGYELDEPAEDDTSNLSFRERRAMGYND-----LPSSLDQALROM 402
Qy 410 KESKMLNKTFCGKILHLYVNAANVEINEFSSKQVTDWELNQGFN 452
Db 403 EKSELVADILGHEVFEFFLRNKRWRDYQEQITPWEELRNLD 445

RESULT 4

US-09-815-242-13378
; Sequence 13378, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13378

LENGTH: 448

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-815-242-13378

Query Match 18.2%; Score 432; DB 9; Length 448;
Best Local Similarity 27.2%; Pred. No. 4.4e-36;

Matches 128; Conservative 83; Mismatches 220; Indels 40; Gaps 10;

Qy 1 MTIYDELNLIRNGKIDTVVLACDQGRMLGRLTGRHFLGDDQ---KKISSTFYVA 57
Db 1 MPITAADIRREVKNKVTIFRLMFSIDILGTMKVNEIP-----ATDEQLDKVLSNKNVMPDG 55
Qy 58 VTIEGIAGGVEISSVDGTGSDCHLCADLNSLHLLPWSE--GAVLAISNPHNFTVTSPLF 115
Db 56 SSISGF-----VRINESDMYLPDLDTWTVPFGDENGVSAGLICDVYTTGEGPFA 106
Qy 116 CSPRVILMQQIERLANLKLGL-FASELEFNLF--NETYKSASOKHWKLNKTAQPHHQM 172
Db 107 GDRGNLKRALRHMEVGFKSNLGPPEFFLFXLDENGDTLLEVNDKG-----GYF 158
Qy 173 NISASSGIETPMRSVRNKLKEAGILMEATHPEFLPSOHELNFVPADPLTMADRHIIAKHG 232
Db 159 DLAPTDLADNTRREIVNVLTKWGPEVEASHHEVAVGQHEIDFKYDEVLRACDKIQIFKL 218

Qy 233 VREMAEQSMVATFMAKLSSTALGNACHIHMSLQDAETKNAFYDONDEYGM--STLARN 290
Db 219 VKTIARKHGLYATFMAKPKFGIAGSGMHCMNSLFDAB-GNNAFPDPNDPKGMQLSETAYH 277
Qy 291 WIAGLLKYVPEATYFFASYINSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNKSEGINVE 350
Db 278 FLAGGLIKHAYNTAIMNPTVNSYKRLVPGYEAPVYIAWAGNRNSPLVRVPASRGMGTRLE 337
Qy 351 LRIGGADLNPLYAFSAIIAAGISGIEBEKLELPPASGNVY-----NDKELPEFPNSLQ 403
Db 338 LRSVDPMANPYVAMAVLLEVLGYLENKIEAPAPIEBENIYMTABEKEKAGITDPLSLH 397
Qy 404 NATHLLKESKMLNKTFCGKILHLYVNAANVEINEFSSKQVTDWELNQGFNRY 454
Db 398 NALKALTEDEVVKAALGDHIYTSFLEAKRIEWASYATFVSQWEIDNVDLY 448

RESULT 5

US-10-156-761-13481
; Sequence 13481, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13481
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-13481

Query Match 18.0%; Score 427.5; DB 15; Length 453;
Best Local Similarity 30.0%; Pred. No. 1.3e-35;

Matches 127; Conservative 74; Mismatches 167; Indels 55; Gaps 12;

Qy 61 EGIAGGGYEISSVDGTG-SDCHLCADLNSLHLLPWSEGA-----VLAISNPHNFTVSEP 113
Db 47 EGIGFDGSAIEGFARVYESDMIAKPPDPSTFQVLPWRAEAGTARMPDCILMP-----DGSP 102
Qy 114 LFCSPRVILMQQIERLANLKLGLFASELEFNLFNETYKSASOKHWKLNKTAQP----- 167
Db 103 SFADPRVYLKRALAKTSDLGFTFTVTHPEIEFFLLKQ-----KPLDGSRRTPADNS 152
Qy 168 ---HHQMNISASSGIETPMRSVRNKLKEAGILMEATHPEFLPSOHELNFVPADPLTMAD 224
Db 153 GYFDHTPQNV---GMD-FRQAITMLESMGISVEFSHHEGAPGQQQIDILRYADALSTAD 207
Qy 225 RHIIAKHGVREMAEQSGMWATFMAKLSSTALGNACHIHMSLQDAETKNAFYDONDEYGM 284
Db 208 NIMTFRLVMQKVALEQQGVQATFMPKPEFSEYPPGSGMTHLSL---FEGDRNAFYEGSGSYQL 265
Qy 285 STLARNWIIAGLLKYVPEATYFFASYINSYKRLQPLT-----FAPTKCCWAIDNRTSA 336
Db 266 SKVGRSFIAGLLKHAABEISAVTNQWNSYKRIWGGSSERTAGAGGAPSYICWGHNNRSAL 325
Qy 337 FRL---CNSKSEGINVELRIGGADLNPLYAFSAIIAAGISGIEBEKLELPPASGNVY--N 391
Db 326 VRVPMYKPGKTGSRVEVRLSDGANPLYALALLAAGLKGIBEGYELPFGADDDVWALS 385

Qy 233 VREMAEQSMVATFMAKLSSTALGNACHIHMSLODAETKNAFYDQNDQVGMSTLARNWI 292
 Db 220 VKTIARKHNLHATFMPKPLFGVNGSGMHFNVL--FKGENAFDPNTEMLGTETAYQFT 277
 Qy 293 AGLKVVPEATYFASVINSYKRLQPLTFAPTKCMAIDNRTSAFLCNSKSEGINVELR 352
 Db 278 AGVLKNARGFTAVCNPLVNSYKRLVPGYEAPCYIANSKGRSPILIRVPSRGLSTRLEVR 337
 Qy 353 IGGADLNPLAFSAIIAAGISGIEKELELPPASGNVY--NDKE-----LPEFPNSLQNA 405
 Db 338 SVDPAANPYMALAAILAAGLDGINKLKVPPEVQNQIYEMNREAREAVGIQDLPSTLYTA 397
 Qy 406 THLLKESKMLNKTGKILHLYNNAANVEINEFSKQVTDWELNOGFNRY 454
 Db 398 LKAMRENEVIKKGALNHIYNQFINSKSIWDYRTQVSEWERDOYMKQY 446

RESULT 8

US-09-815-242-12775
 ; Sequence 12775, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 12775

; LENGTH: 446

; TYPE: PRN

; ORGANISM: Staphylococcus aureus

US-09-815-242-12775

Query Match 17.9%; Score 424.5; DB 9; Length 446;
 Best Local Similarity 26.2%; Pred. No. 2.7e-35;
 Matches 123; Conservative 92; Mismatches 211; Indels 43; Gaps 9;

Qy 2 TITVDELNLIRNGKIDTVVACVDMQGRMLGKRLTGRHFLGLDQKISISTFVYAVTIE 61

Db 5 TFTKDDIRKAEENVYRLQFTDILGTIKNVEVPVSQL-----EKVLDMEMFDDGSIE 60

Qy 62 GIAGGGVEISVDTGSDCHLCADLNSHLPLWS--EGAVLAISNPHNFVTSEPLFCSPR 119

Db 61 GF-----VRIESDMVLHPDLDTWIFPWTAGQGVKVARLICDVYKTDGTFEGDPR 111

Qy 120 VILMQOIERLANLKLGL--FASELEFNLFNETHYKASQKHWKLUKTAQPHHQ-----WM 172

Db 112 ANLKRVLKEMEDLGTDFNLGPEPEFFLKLDEK-----GEPTLELNDGGYF 159

Qy 173 NISASSGIETFMRSVRNKLSEAGILMEATHPEFLPSQHELNFPVADPLTWADRHIIAKHG 232
 Db 160 DLAPTDLGENCRDILELEDMGDFIEASHHEVAPGQHEIDFKYADAVTACDNIQTQKLV 219
 Qy 233 VREMAEQSMVATFMAKLSSTALGNACHIHMSLODAETKNAFYDQNDQVGMSTLARNWI 292
 Db 220 VKTIARKHNLHATFMPKPLFGVNGSGMHFNVL--FKGENAFDPNTEMLGTETAYQFT 277
 Qy 293 AGLKVVPEATYFASVINSYKRLQPLTFAPTKCMAIDNRTSAFLCNSKSEGINVELR 352
 Db 278 AGVLKNARGFTAVCNPLVNSYKRLVPGYEAPCYIANSKGRSPILIRVPSRGLSTRLEVR 337
 Qy 353 IGGADLNPLAFSAIIAAGISGIEKELELPPASGNVY--NDKE-----LPEFPNSLQNA 405
 Db 338 SVDPAANPYMALAAILAAGLDGINKLKVPPEVQNQIYEMNREAREAVGIQDLPSTLYTA 397
 Qy 406 THLLKESKMLNKTGKILHLYNNAANVEINEFSKQVTDWELNOGFNRY 454
 Db 398 LKAMRENEVIKKGALNHIYNQFINSKSIWDYRTQVSEWERDOYMKQY 446

RESULT 9

US-09-815-242-10514
 ; Sequence 10514, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 10514

; LENGTH: 451

; TYPE: PRN

; ORGANISM: Enterococcus faecalis

US-09-815-242-10514

Query Match 16.9%; Score 402; DB 9; Length 451;
 Best Local Similarity 28.9%; Pred. No. 6.1e-13;
 Matches 114; Conservative 63; Mismatches 184; Indels 34; Gaps 7;

Qy 78 SDCHLCADLNSHLPLWS--AVLAISNPHNFVTSEPLFCSPRILMQOIERLAN 131

Db 73 SDWYLPDLSTWIFPWSHDGKVARLICDINP---DGTTFAGDR-----GN 118

Qy 132 LKUKGLFASLEFNLFN-----ETYSASQKHWKLUKTAQPHQNMNISASSGIETPMRS 186

Db 119 LKRALADMKBELGTSFNLGPEPEFFLKLDEGEITDLDNDKGGYDFDPTDLGNCRRD 178

QY 187 VRNKEEAGILMEATPELPSOHELNFVPAADLTWADRHIIAKHGVRMAQSGMVATF 246
 Db 179 IVLESLGFEVASHVAPGHEIDFKYADVWDACDNIOTFKLVVKTIAKHGLHATF 238
 QY 247 MAKLSGALGNACHIHMSLODAETKNAFYDQNDYGMSTLARNWTAGLLKXVPEATYFF 306
 Db 239 MPKPLFGINGSGHCHMSL--FNEEGNVFDESSEGLSOTAYHFLGGLKHARAYTAVC 296
 QY 307 ASYINSYKLOPLTFAPTCKCWAIDNRTSAFLCNKSGSINVELRIGGADLNPLAPSA 366
 Db 297 NPTVNSYKRLVPGYEAPVTVVAMSGRNSPLIRVPESRGLSTRLELRSDVPSANPLYTMAV 356
 QY 367 ITAAGISGIEEKLLEPPASGNVY--NDXE-----LPEPPNSIQNATHLLKESKMLNKTFF 419
 Db 357 LIQAGLDGRNELTPPPADVDRNIYVNEERQHAQIEDLPSTLHNAIKELRQDQWIDAL 416
 QY 420 GEKLILHYVNAANVEINEFSKVQTDWELNQGNRY 454
 Db 417 GRHIFANFVEAKRMWAAFRQTVSEWEREQYLELY 451

RESULT 10
 US-09-738-626-6985
 ; Sequence 6985, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent In ver. 3.0
 ; SEQ ID NO 6985
 ; LENGTH: 477
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6985

Query Match 13.6%; Score 322; DB 10; Length 477;
 Best Local Similarity 26.6%; Pred. No. 1.6e-24;
 Matches 124; Conservative 74; Mismatches 207; Indels 62; Gaps 14;
 QY 4 TYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGI 63
 Db 5 TPEIVKFIKDNVEFVDRFTDLPG-----TEQHF-----SIPAAAFDADTIEGL 51
 QY 64 AGGYEISSVDTG--GYSDCHCADLNSLHLLPWSGAVLAIS--NPHNFVTSPLFCSPRV 121
 Db 52 AFDGSSIRGFTTIDSDMNLPLDGTATLDPFRKAKTLNVKFFVHDPPFTREAFSRDPRNV 111
 QY 122 LMOQIERLANLKUGL--FASELEFNLNET-----YKSASOKHWNKLTQAQPH 169
 Db 112 ARKAEQYLASTGADTCNFGAEAEFYLFDSVRYSTWNSGFEYVDTEBGMWNRGKTNLD 171
 QY 170 QMWNISASSGIE-----TFMRS--VANKLEEAGILMEATHPEFLPSQHELNFV 215
 Db 172 GTNLAGKRNKVGKGYFFVAPYDQTVDRDDMVN--LAASGAFALRPFHFHEVGQOQBINYR 230

QY 216 PADPLTMADRHIIAKHGVRMAEQSGMVATFMAKLSSTALGNACHIHMSLODAETKNAF 275
 Db 231 FNTMLHAADDQTQFKYIKNTARLHGKAATFMPKPLAGDNGSGMHAQSLW--KDGKPLP 288
 QY 276 YDQNDYGMSTLARNWTAGLLKXVPEATYFFASYINSYKRLQPLTFAPTKCWAIDNRTS 335
 Db 289 HDESGVAGLSDIARYIIGILHAGAVLAFNTATLNSYHRLVPGPEAPNLVYSQRNSA 348
 QY 336 APR--LCNSKSEGINVELRIGGADLNPLAPSAIIAAGISGIEEKLLEPPASGNVY--- 390
 Db 349 AVRIPITGSPKAKRIEPRAPDPSGNPYLGFAMMMAGLDGKRIEHPHAPVDKOLYELP 408
 QY 391 --NDKLEPFPNS-----LQNATHLLKESKMLNKTFFGEKLILHYV 428
 Db 409 PEEAASIFQAPTSLEASLKALQEDTDFLTESDV-----FTEDLIEAYI 451

RESULT 11
 US-10-156-761-13532
 ; Sequence 13532, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 13532
 ; LENGTH: 469
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-13532

Query Match 12.2%; Score 291; DB 15; Length 469;
 Best Local Similarity 25.9%; Pred. No. 2.8e-21;
 Matches 113; Conservative 68; Mismatches 189; Indels 66; Gaps 13;
 QY 6 DELANNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFVYAVTIEGIAG 65
 Db 6 DEAKYIADDEVKFDVVRFCDDLPFGV-----QHF-----TIPAEAFDPA---EELAF 49
 QY 66 GGYEISSVDTGY--SDCHCADLNSLHLLPWSGAVLAIS--NPHNFVTSPLFCSPRVILM 123
 Db 50 DSSIRGQAIHESDMALRADLSTARVDFPRDKTVNINFFIHDPTGQYSRDPENVAK 109
 QY 124 QOIERLANUKL--KGLFASELEFNLNET--YKSASOKHWNKLT----- 164
 Db 110 KAEAYLASTGIADTAYFGPEAEFYFVDSVRFKTAENESFYHIDSEAGAWNTGAIEDNRY 169
 QY 165 -----AOPHHQWNISASSGIEETFMRSVRNKLKEEAGILMEATHPEP--LPSQHEIN 213
 Db 170 KVRVYGGYFPTTPVDHFDLRAEISLE-----LANSGLQVERQHHVEVGTAGAEIN 220
 QY 214 FVPADPLTMADRHIIAKHGVRMAEQSGMVATFMAKLSSTALGNACHIHMSLODAETKN 273
 Db 221 YKFNILLAAADDQLFKYIKVKNVNRNGKTAFTFMPKPIFGDNGSGMWHVHOSLW--ANGDP 278
 QY 274 AYDQNDYGMSTLARNWTAGLLKXVPEATYFFASYINSYKRLQPLTFAPTKCWAIDNR 333
 Db 279 LFYDEAGYAGLSDTARYIYGILKHAQSLAFNTATLNSYHRLVPGFEAPVNMVYSQRNR 338

Qy 334 TSAPR--LCNSKSEGINVELRIGADLNPNYLAFAIAAGISGIEEKLLEPPASGVNY- 390
Db 339 SAAMRIPITGSPKAKRVEFRAPDSGNPNYLAFAISALLAGLDGINKKVEPAEPIDKOLYE 398
Qy 391 ----NDKELPEFPNSL 402
Db 399 LAPEHASVAQVPTSL 414

RESULT 12

US-09-815-242-11137
; Sequence 11137, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11137
; LENGTH: 472
; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-815-242-11137

Query Match 12.1%; Score 287; DB 9; Length 472;

Best Local Similarity 24.3%; Pred. No. 7.3e-21;

Matches 116; Conservative 91; Mismatches 192; Indels 78; Gaps 19;

Qy 11 LIIRNGKIDTVVLACVDMQGRMLGKRLTGRIHFLGLDQKKISISTVYAVTIEGAGGYE- 69

Db 13 LIENNKKVLLRFTDIKK-----EHGVSIPVSLVDEMDPEDGKMFDSGSSVEGKMT 64

Qy 70 ISSVDTGYSCHLCLADLNSLHLLPWSRGVLAISNPHNFTVSEPLFCS----- 117

Db 65 INKAD-----MLLMPMAETAIIV---DPFAQPTLSIRCSVVEPTMQSYDR 107

Qy 118 -PRVILM--QQIERLANLKLGLFASLEFNLFNETYKSA-----QKHWK-NL 162

Db 108 DPRSIAIRAENYMRSTGIAQAFGPPEPFPLDPDVRFNYSMNKASFSIDIEAAMNTK 167

Qy 163 KTAQPHQWNNIS-----ASSGIETFMRSVRNK-----LEBAGILMEATHPEFLPSQHELN 213

Db 168 KYEEGNNAVPLKKGKGCYCAVAPIDS-AHDIRSEMCLILEEMGLVIEAHHHEVATAGONEI 226

Qy 214 FVPADPLTM-ADRHIIAKHGVRMAEQSGMVATFMKLSSTALGNACHIIHMSLODAETEK 272

Db 227 ATKFNLTUTLKADETQIIYKHVVQNVVALEHGKTACMPKPTIGDNGSGMHCMNSL--SKDG 284

Qy 273 NAFVDQNDY- GMSITLARNIAGLLKXVPEATYVFASYINSYKRLQLPTFTAPTCCWMAID 331
Db 285 NIF--QGDKYAGLSETALYVIGGIIKIAKALNAFTNPSTNSYKRLVPGYEAPVLLAYSAS 342

Qy 332 NRTSAFEL-CNSKSEGINVELRIGADLNPNYLAFAIAAGISGIEEKLLEPPASGVNY 390

Db 343 NRSASIRIPAVTNPKAIRVEARFPDPLANPYLAFAALLMAGLDGVNKHIFPGDAMDKNLY 402

Qy 391 N-----DKELPEFPNSLQNAHLI-KESKMLNK--TFGEKLIILHYVNAANVEINEFS 439

Db 403 DLPPEELKDIPAVASSILEEALNSLEKDYEFYLTQGGVFAKOFIDAFISIKRKEVERLN 459

RESULT 13

US-10-187-267A-57

; Sequence 57, Application US/10187267A

; Publication No. US20030124679A1

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; APPLICANT: Paradkar, Ashish

; APPLICANT: Varoglu, Mustafa

; APPLICANT: Mathur, Eric J.

; TITLE OF INVENTION: GLYCOSYLATED KINAMYCINS AND METHODS OF

; TITLE OF INVENTION: MAKING AND USING THEM

; FILE REFERENCE: 09010-280001

; CURRENT APPLICATION NUMBER: US/10/187.267A

; CURRENT FILING DATE: 2003-01-27

; PRIOR APPLICATION NUMBER: US 60/301,401

; PRIOR FILING DATE: 2001-06-27

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 57

; LENGTH: 500

; TYPE: PRT

; ORGANISM: Streptomyces murayamaensis ATCC 21414

US-10-187-267A-57

Query Match 11.8%; Score 281.5; DB 15; Length 500;

Best Local Similarity 25.1%; Pred. No. 3e-20;

Matches 122; Conservative 77; Mismatches 210; Indels 77; Gaps 20;

Qy 16 KIDTVVLACVDMQGRMLGKRLTGRIHFL-----GLDQKKISISTVY-----AVTIEGIAGG 66

Db 45 EIDLVLRAFADPHGLARSKTLTVDAFRSLVRNGMD---FSSGPFIFDTGHAALDFLADP 101

Qy 67 GYEISSVDTGYSCHLCLADLNSLHLLP-----WSEG-AVLAISNPHNFTVSEPLFCS 118

Db 102 GVGVDI-AGAGSFVLVPDPLTFQVLPHGEGRTAWVLGDBYLRDGSHP-----PL--SA 152

Qy 119 RVILMQOIERLANLKLGLFASLEFNLFNETYKSA-----SQKHWNKLTQAPHHWNNI 174

Db 153 RNVLRQVARYAARDLTPVLGLEVWVL---TRKLACPPGNAGNGFLOGAAPKVEAMNS 209

Qy 175 SASGIETFMRSVRN-----KLEBAGILMEATHPEFLPSQHELNFPVADPLTMADRI 227

Db 210 GYOFNLNDANYDSVAHFTSPLAMKLELGLPLRSIEHSGPGQIETTFNPMHALDTADAM 269

Qy 228 IAKHGVRMAEQSGMVATFMKLSSTALGNAC--HIHMSLODAETEKNAFYDQNDYCMS 285

Db 270 LFTVTVKQTAAROGYHASFMA-LPRVDSFDPCCWHHLQSVMDSTNGRNL- -AADGGGIS 326

Qy 286 TLARNIAGLLKXVPEATYVFASYINSYKRL-QPLTFAPTCCWMAIDNRTSAPRLCNSKS 344

Db 327 DQKAYIDGLLSRARDLCLLSVFTNGYRMGADFSLSPTRLGWSVEDRAMIRVVCG-G 385

Qy 345 EGINVELRIGADLNPNYLAFAIAAGISGIEEKLLEPPASGVNYNDKELPFPPNSLQN 404

Db 386 AGTHIENRVGEPTANPYLNIAAQLSAGFDGMATEVGASTRESG-----EESVETLPQDLGE 441

Qy 405 ATHLLK---ESKMLNTFGEKLIILHYVNAANVEINEFSK-----QVTDWELN 448

Db 442 ALEAFRAGQAELLGKPLAATL-----AKKSELSRYEAWRAAERPADGOVTWEQOR 494

```
QY 449 QGFNRY 454
Db 495 EYFEAF 500

RESULT 14
US-09-815-242-12088
; Sequence 12088, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12088
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12088

Query Match 11.2%; Score 265.5; DB 9; Length 469;
Best Local Similarity 24.8%; Pred. No. 1.3e-18;
Matches 119; Conservative 79; Mismatches 194; Indels 87; Gaps 20;

QY 9 NNLIRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISITFTFYAVTIEGIAGGY 68
Db 6 HQLIKDHVDKVDLRTDTYKGQOQVTPARD--ALDDREFF-----AGKMF 50

QY 69 EISSVDGTGSDCHLCADLNSLHLLPWSSEGAVLAISNPHNFVTSEPLFCSPRVIL----- 122
Db 51 DGSSI-AGWKG-----IEASDMILMPDDSTAVL-----DPTEETLLIVCDIIE 94

QY 123 ---MQQIER---LAN-----LKLGL-----PASELEFNLFNET-YKS-----A 154
Db 95 PSTMQGYERDPRNIAKRAEYLSGTGDTVFVGPEPEFFIDFVFKSDISGSMFKIFS 154

QY 155 SQQHWK--NLKTAQPHQ-----WNISASSGIETFMRSVRNKLKEEAGILMEATHPEF 205
Db 155 EQASWNTDIESGNGKRGPGVKGYPVPVDDHDEIRTMCNALKEEGLVVEVHHHEV 214

QY 206 LPS-QHELFNVPADPLTMADRHIIAKHGVREMAEQSGWATFMAKLSSTALGNACHIMS 264
Db 215 ATAGQNEIGVKFTLVAKADVQTLKCVHNVADVGKTVTFPKPLYGDNDSGMVHMS 274

QY 265 LQDAETEKAFYQNDDEYGMSTLARNWIAGLLKYVPEATYFFASYNSYKRLQPLTFAPT 324
Db 275 I--SKDGKNTFAGE--YAGLSETALYFIGGIIKHGKALNGFTNPSTNSYKRLVPGPEAPV 331

QY 325 KCCWAIDNRTSAFRL-CNSKSEGINVELRIGSGADLNPYLAFSAIIAAGISGIEEKLLEPP 383
Db 332 MLAYSARNRSASIRIPYVUSPKARRIEARFPDPAANPYLAFAPALMAGLDGIONKIHPGD 391

QY 384 PASGNVYN-----DKELPEFPNSLQATHLLKESK-MLNK--TFGEKLIHLHYVNAANVE 434
Db 392 AADKNLYLDPPEAKEIPQVCGSLKEALELDKGRAFLTKGGVFTDEFIDAYIELKSEE 450

RESULT 15
US-09-815-242-11507
; Sequence 11507, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11507
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11507

Query Match 11.2%; Score 265; DB 9; Length 481;
Best Local Similarity 24.7%; Pred. No. 1.5e-18;
Matches 100; Conservative 80; Mismatches 181; Indels 44; Gaps 13;

QY 77 YSDCHLCADLNSLHLLPWS-EGAVLAISNPHNFVTSEPLFCSPRVILMQOIRLANLKLK 135
Db 72 HSDMLTDPDLVRYFIDPFSDVSAVVFCVDYDVYKNQPYEKPCPSIAKALKALHRLDLGLG 131

QY 136 GL--FASELEFNLFNE-TYKSAS-----QKHWNKLTQAQ-----PHHQ--WM 172
Db 132 DVAYFGAENEFIFDPSIKIKDASNSQYVEVDSEGEWNDRKSFENGWVNFHPRGQGGYM 191

QY 173 NISASSGIETFMRSVRNKLKEEAGILMEATHPEFSPQHELFNVPADPLTMADRHIIAKHG 232
Db 192 PVPPTDITMDIRTEIVKVLNVGLTFVHVHVEVAQAQGEVGVKFGDLVEADNVOKLKYV 251

QY 233 VREMAEQSGWATFMAKLSSTALGNACHIMSLQDAETEKNAFYQNDDEYGMSTLARNWI 292
Db 252 VKWVAHLNGKTAFTMPKPLYGDNDSGMVHVSIGW--KNENLFGGETYK-GLSEFALYEL 308

QY 293 AGLLKYVPEATYFFASYNSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSE-GINVEL 351
Db 309 GGVLRHARGLAFTNASTNSYKRLPGVEAPSLITYSASNRSAVRIPYGISKNSARFEP 368
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:43:05 ; Search time 19 seconds
(without alignments)
1011.007 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MTITYDELNLIRNGKIDTV.....INEFSKQVTDWELNGQFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCFUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	20.2	449	4	US-09-252-991A-17901
2	459	19.3	464	4	US-09-252-991A-27559
3	455	19.1	464	4	US-09-252-991A-28048
4	450	18.9	489	4	US-09-252-991A-27833
5	409	17.2	452	4	US-09-134-001C-4173
6	389	16.4	531	4	US-09-252-991A-26690
7	344.5	14.5	491	4	US-09-252-991A-19014
8	311.5	14.0	480	4	US-09-328-352-6949
9	317	13.3	1037	4	US-09-252-991A-16796
10	280.5	11.8	661	4	US-09-252-991A-18960
11	265.5	11.2	481	4	US-09-252-991A-32811
12	255.5	10.8	489	4	US-09-328-352-4363
13	241.5	10.2	190	4	US-09-107-532A-4613
14	134	5.6	275	4	US-09-107-532A-6200
15	95	4.0	481	4	US-09-328-352-4580
16	95	4.0	1338	1	US-08-471-033-50
17	95	4.0	1338	2	US-08-471-044-50
18	95	4.0	1338	2	US-08-463-483A-50
19	95	4.0	1338	2	US-08-471-046A-50
20	95	4.0	1338	2	US-08-470-566B-50
21	95	4.0	1338	2	US-08-469-334-50
22	95	4.0	1338	3	US-09-300-529-50
23	94	4.0	291	2	US-08-701-191A-24
24	94	4.0	606	3	US-08-335-865J-20
25	92	3.9	527	3	US-09-028-366-2
26	92	3.9	527	3	US-09-028-366-3
27	91.5	3.9	822	4	US-09-252-991A-29534

28	91	3.8	461	4	US-09-134-001C-5311
29	91	3.8	593	3	US-08-335-865J-21
30	91	3.8	1013	3	US-09-415-522-8
31	90	3.8	566	3	US-08-335-865J-9
32	90	3.8	10182	4	US-09-134-001C-3159
33	89.5	3.8	506	4	US-09-134-001C-4383
34	89.5	3.8	952	4	US-09-328-352-4551
35	88.5	3.7	731	4	US-09-252-991A-18769
36	88	3.7	2152	3	US-09-036-987A-3
37	88	3.7	2152	3	US-09-370-700-3
38	88	3.7	2152	4	US-09-603-207-3
39	87.5	3.7	441	4	US-09-254-776B-79
40	87	3.7	299	4	US-09-328-352-7993
41	87	3.7	733	4	US-09-328-352-5599
42	87	3.7	1074	1	US-07-912-015-4
43	86.5	3.6	605	3	US-09-063-950-5
44	86.5	3.6	3290	4	US-09-328-352-5486
45	86	3.6	973	4	US-09-107-532A-4810

ALIGNMENTS

RESULT 1
US-09-252-991A-17901
; Sequence 17901, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17901
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17901

Query Match	20.2%	Score 481	DB 4	Length 449
Best Local Similarity	28.9%	Pred. No. 2.2e-44		
Matches	128	Conservative 94	Mismatches 181	Indels 40
Gaps	11			
QY	22	LACV--DMOQRLMGKRLTGRHFLGLDQDKKISITGFVYAVTIEGIAGGYEISSVDTGYSD	79	
Db	21	IECVTPDLNGVPRGKWTGEGF--LSGRRLQLAR---GVLLQIMGYPPARFVGSDDG	75	
QY	80	CHLCADLNSLHLPSWSEGAVALISNPHNFTSE-----PLFCSPRVILMQOIERLANL	132	
Db	76	LALVAEPTQVHRLPWS-----NTPRAFAICDAQELDGTGSLSTRGLLKQVVARVAH	128	
QY	133	KLKGLFASELEFNLF-----NETYKSASQKHWNKLTQAPHHQWNISASSGIETPMRS	186	
Db	129	GWQFVVAATELEFFVFPAPNTDPNEPQAPLPGDGR---ELGYSAFSVSSNGLRPPFD	184	
QY	187	VRNKLKEAGFLMEATHEPLPSQHELNFFVADPLTWADRHIIAKHGVRMAEQSGMVATF	246	
Db	185	VYRCMDALGLVURTDFHEMGTSOFENFLHGDVPVLADQTFLEKLLKVALKHGLIIVC	244	
QY	247	MAKLSSTALGNACHIHMSLQDAETEKNAFYDQNDXYGMSTLARNWLAGLLKYVEATYFF	306	
Db	245	MAKPLAKTPGSSMHIIHQSIIVELDGGRIIFSEADGE--PSAAFRHFIGGQACIADFTFF	302	
QY	307	ASTVNSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNKSEGINVELRIGGADLNPYLAFA	366	
Db	303	APHVNSFORLCHPYSPNNAWNSHDNRAGLRIPASGASARRVENRPLGDANPYLAIAA	362	

Qy	367	IINAAGISGIEEKLELP PPPASGNVYNDKELPE---FPNSLQNATHLLKESQVLNKTGEKL	423
Dd	363	SLAAGLYGLEHELE-PSPA---IQGFVPPELTLPCTMYDALRRLKGSALARELFGSEF	418
Qy	424	ILHYNNAANVEINEFSKVQTWDE	446
Dd	419	VEGYVATKSMELTSPFFDEISPWE	441

RESULTS

```

RESULI 2
US-09-252-991A-27559
; Sequence 27559, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27559
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27559

```

RESULT 3

US-09-252-991A-28048
; Sequence 28048, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27833

Query Match 18.9%; Score 450; DB 4; Length 489;
Best Local Similarity 28.4%; Pred. No. 7.3e-41;
Matches 130; Conservative 80; Mismatches 222; Indels 26; Gaps 7;

```
Qy 1 MTITYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGDLQDKKISISTFVYAVTI 60
Db 38 MTTKLDQLTSLWKERTITEVECLISDLTGARGKISPTNKFAISKGRMLPESVLLQIVTG 97
Qy 61 EGIAGGGYEISVDTGSDCHLCLADLNSLHLLPWS-EGAVLAISNPHNFTVSEPLFCSPR 119
Dy 98 DYVEDDIY-YDLLDPADIMVCRDENAVFLVPAIEPTAMVTHDTDKL-GNPDIELSPR 155
Qy 120 VILMQOIERLANLKLGLFASELEFNLFNET-----YKSASQKHWNKLTQAQPH 168
Dy 156 NILKRVLNKADKGRWRIIVAFEMEFYLTKRSDDPDPLQAPVGRGQ-----ETG 206
Qy 169 HQWMNISASSGIETFMRSVRNKLKEAGILMEATHPEFLPSQHELFNFPADPLTMADRHII 228
Dy 207 RQSFIDAANEFDFLEDMDYDWCQAQGLDLTLTHEEGTAQWEINFRHGDALDLADQILV 266
Qy 229 AKHGVREMAQSGMVATFMAKLSSTALGNACHIIHMSLODAETKNAFYDQNDYEGMSTLA 288
Dy 267 FKRTMREAALKHNVAAATFMKPMTCGPGSAMHLHQSIYDVVKTKGNIF--SNADGTMSELF 324
Qy 289 RNWITAGLLKYVPEATYFASVINSYKRLQPLTAPTCKCWAIDNRTSAFRLCNKSKSEGIN 348
Dy 325 LHHIGGLOKFIPEVLPLFAPNVNSFRPLDPTSAFNVNVEGEENRTVGLRVPDSSPENRR 384
Qy 349 VELRIGGADLNPLYLAFSAIIAAGISGIEEKLPPASGNVYNDKELPEFPNSLQONATHL 408
Dy 385 VENRLAGADANPLYLAAASLLCGYIGVVEGKPSAQVKGGRYERNL-RLPLTIEAALER 443
Qy 409 LKSKMLNKTGFEKLIHLYNNAANVEINEFSKQVTDWE 446
Dy 444 MENCKPLEQYLGSKFISGYAVKRAEHENFKRVISSWE 481
```

RESULT 5
US-09-134-001C-4173
; Sequence 4173, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134.001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4173
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4173

Query Match 17.2%; Score 409; DB 4; Length 452;
Best Local Similarity 28.2%; Pred. No. 2.3e-36;
Matches 111; Conservative 77; Mismatches 175; Indels 30; Gaps 7;

```
Qy 78 SDCHLCLADLNSLHLLPWS--EGAVLAISNPHNFTVSEPLFCSPRVILMQOIERLANLKLK 135
Dy 74 SDMYLHPDLDTWVIFPWTAGQKVARLICDVKTDGTFPEGDPRLNKLKRLRMEDGFT 133
Qy 136 GL-FASELELRLNFTNETYKSAQKHWNKLTQAQPHQ-----WNNISASSGIETFMRSVR 188
Dy 134 DFNLGPEPEFFLKLDEK-----GEPTLELNDGDPDLPATLGENCRDRIV 181
```

```
Qy 189 NKLEEAGILMEATHPEFLPSQHELFNFPADPLTMADRHIIAKHGVREMAQSGMVATFMA 248
Dy 182 LELEDWGFDEASHHEVAFQCHIEDFKYADAVTACDNIOTFKLVVTKTIARKHNLHATFMP 241
Qy 249 KLSSTALGNACHIIHMSLODAETKNAFYDQNDYEGMSTLARNIAGLLKLVPEATYFAS 308
Dy 242 KPLFGVNGSGMHFNVSLL--FKGKENAFDFDEGDLQTLDTAYQFTAGVLKNARGFTAVCNP 299
Qy 309 YINSYKRLQPLTAPTCKCWAIDNRTSAFRLCNKSKSEGINVELRIGGADLNPLYLAFSAII 368
Dy 300 IVNSYKRLVPGYEAPCYIAWSGKNRSLVRVPTSRGLSTRIEVRSVDPAANPNYALAAAIL 359
Qy 369 AAGISGIEEKLPPASGNVY--NCKE-----LPEPPNSLQONATHLLESKMLNKTGFE 421
Dy 360 EAGLDGIENKLEVPENPVNQNIYENREAREAVGIQDLPLSTLYLTALKAMRENKSIKNALGN 419
Qy 422 KLILHYNNAANVEINEFSKQVTDWELNQGNRY 454
Dy 420 HIYNQFINSKSIEWDYVRTQVSEWEREQYIKQY 452
```

RESULT 6
US-09-252-991A-26690
; Sequence 26690, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26690
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26690

Query Match 16.4%; Score 389; DB 4; Length 531;
Best Local Similarity 27.6%; Pred. No. 5.1e-34;
Matches 132; Conservative 66; Mismatches 208; Indels 72; Gaps 13;

```
Qy 6 DELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGDLQDKKISISTFVYAVTIEGIAG 65
Dy 87 EELKQQLQDKGVKYMASVYDIHGVIKGFVPIAH-LGQMLRGSLEYT----- 133
Qy 66 GGYEISSVDTGYSDCHLCA--DLNSLHLLPWSGAVLAISNPHNFTVSEPLFCSPRVILM 123
Dy 134 -GAALDGVPOEISDNEVAAMPDPATATQCSWNRD--LAWPASDLYLDGKFPDACSRLGILK 190
Qy 124 QOJERLANLKLGLFASELEFNLFNET-----YKSASQKH-----WNKLTQAQ 167
Dy 191 RQTEAAALGYTFTNLGIETETFEFLFKDTPDGGFAPISERDDMAKPCVDPRLMLNPLIVDE 250
Qy 168 HHQWMNISASSGIETFMRSVRNKLKEAGILMEATHPEFLPSQHELFNFPADPLTMADRHII 227
Dy 251 LVQAMN-EMGWVYSFDHEDAN-----GQETDFKYTDALGMADRFV 291
Qy 228 IAKHGVREMAQSGMVATFMAKLSSTALGNACHIIHMSLODAETKNAF-YDQNDY--GM 284
Dy 292 FFFMMAEIAARKHGAETFMKPSANRTSGAHYNNSLADIETGKNLFEVDGEDPYCGGV 351
Qy 285 STLARWNIAGLLKLVPEATYFASVINSYKRL-----QPLTAPTCKCWAIDNRTSAFR 338
Dy 352 TPLAYHFIAIGVLKHAKAICAVIAPTYSYKRLIRKGMASGSTWAPVCCYGNRRNMLR 411
Qy 339 LCNSKSEGINVELRIGGADLNPLYLAFSAIIAAGISGIEEKLPPASGNVYNDKE----- 394
```

Db 412 I---PQGARVECAADICGPNYPYLAAGLEGIRKLDQPGQPHREMYHSREVA 468
Qy 395 ----LPFPNSLQWATHLLKESKMLNKTGFKLILHYVNAANVEINFEFSKVQVTDWELNQ 449
Db 469 QMGIELPRTLSEDAIDAFADPLSLRQVFGDAMYQAFVFKRDEWNAHYTHVSDWEIQ 526

RESULT 7

US-09-252-991A-19014
; Sequence 19014, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19014
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19014

Query Match 14.5%; Score 344.5; DB 4; Length 491;
Best Local Similarity 26.8%; Pred. No. 4e-29;
Matches 122; Conservative 73; Mismatches 231; Indels 29; Gaps 11;
Qy 11 LIRNGKIDTVLVACVDMQGMKRLTGRHFLGLDQKKISISTFVAVTIEGAGGYEI 70
Db 49 LAAHPEVRSIELFLIDANGVPRGKLLHRLDELLAVYESGRPLPSTILGLTIN--GDDVEE 105
Qy 71 SSV--DTGVSDCHLCADLNSLHLLPWSE---GAVLAISNPHNFTVSEPLFCSPRVILMQ 125
Db 106 TGLVWDVGADCWTFPLPGSLTLQPRQSPGTQGVQVSMHPGLGPA--GDRHVLQRV 163
Qy 126 IERLANLKLGLPASELEFNLFNETYKS-----ASQKHWNKLTQAOPHHQWNNISASSG 179
Db 164 IHSLOAEGHPVNAVELEFLDRDSDGRPLPALQMGQRPATQVY---GVYELQ 219
Qy 180 IETFMRSVRNKLEBEAGLMEATHPEPLPSQHELNFPVA-DPLTMADRHIIAKHGVREMAE 238
Db 220 LQPFLLDLVAACEAQLPARTAISVAPGQVEITLQHRFDTLQAIDEGVRYKRLKGVAN 279
Qy 239 QSGWVATFWAKLSSTALGNACHTHMSLQDAETEKNAFYQNDYCGMSTLARNWIALGLKY 298
Db 280 RHGLQACFWAKPPADLSGGLHLVSLADA--AGNNLFASEDPA--TPLLQAIQGMKAC 336
Qy 299 VPEATYFFASYINRYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGCADL 358
Db 337 LLESALFCPNANSPRFOANSVAPLAPTGWGNNRTVSLRVPCCPASSHIEHICGADA 396
Qy 359 NPYLAFSAITAGISGIEEKLPELPPASGNVYNDKELPEFPNSLQWATHLLKESKMLNKT 418
Db 397 NPYLAAAALLAAVRLGIRERLDGPATITNGY--AQATQALPSDWLTALRALEGSAAWREA 455
Qy 419 FGEKLILHYVNAANVEINFEFSKVQVTD-----WELNQ 449
Db 456 LGEDFLKIYLAIKQAEYRAFMEVGEQDWRWYLNQ 490

RESULT 8

US-09-328-352-6949
; Sequence 6949, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6949
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6949

Query Match 14.0%; Score 331.5; DB 4; Length 480;
Best Local Similarity 25.7%; Pred. No. 1.1e-27;
Matches 117; Conservative 76; Mismatches 187; Indels 75; Gaps 14;
Qy 26 DMQGRMLGKRLTGRHFLGLDQKKISISTF---VTVAVTIEG-----IAGGYEISSVDT-- 75
Db 68 DLNGHIRGKRIDVKSL-----KNLSNGCVFPLSVYVAMSLDGKVIETGLGKYIGEPDRLC 122
Qy 76 ----GYSDCHLCADLNSLHLLPWSEGAVALAISNPHNFTVSEPLFC--SPRVLMOQIER 128
Db 123 LPIGLSLOPSALSPELNAQLYLSMOE-----EDSGCRYEPRNLIKLLNQ 168
Qy 129 LANLKLGLFASELEFNLFNETYKSASQKHWNKLTQAOPHHQWNNISASSGIETFMRSYR 188
Db 169 LHANNYFPVMAAEFLYLFSPQHS-----EISCENOCFDDIDANNYQ---QVL 214
Qy 189 NKLEBAGIL---MEATHPEPLPSQHELNFPV-PADPLTMADRHIIAKHGVREMAOSGMV 243
Db 215 DEVERAALQSTIEITAIVAESSPGQVELNLQNHDLKCDQINALKRIVKQVARKHDLT 274
Qy 244 ATFMAKLSSTALGNACHTHMSLQDAETEKNAFYQNDYCGMSTLARNWIALGLKYVPEAT 303
Db 275 ACFMAKPNLAKASGGMHFMSLN-QYHONIFSSSEKDEKALKLSAISGLIELMPASM 333
Qy 304 YPFASYINRYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGGADLNPYLA 363
Db 334 ALAPNINSYRRFKIGHVPLEANWDNTRNVAIRIPCSDVQNRLEYRVAGADCNPLYLV 393
Qy 364 FSAITAGISGIEEKLPELPPASGNVYNDKELPEFP-----NSLQWATHLLKESKMLN 416
Db 394 TATILAGASYGLSHKLLPLPKA-----HLLKFPDEHILLANNQPEALKIFKGSLLIK 445
Qy 417 KTFGEKLILHYVNAANVEINFEFSKVQVTD-----WEL 447
Db 446 GYLGAADFVEHWTYVQAEYQNIYSQMTAEQBHWDI 480

RESULT 9

US-09-252-991A-16796
; Sequence 16796, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16796
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (769)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

Db 398 AADKNLYDLPPEAKEIPQVCGSLKEALBELDKGRAFTKGGVTFDEFIDAYIELKSEE 456

RESULT 12

US-09-328-352-4363

; Sequence 4363, Application US/09328352

; Patent No. 6582958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 4363

; LENGTH: 489

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-4363

Query Match 10.8%; Score 255.5; DB 4; Length 489;

Best Local Similarity 24.7%; Pred. No. 3.1e-19;

Matches 98; Conservative 57; Mismatches 156; Indels 85; Gaps 12;

QY 112 EPLFCSPRVLIM-----QQIERLANLKLK-----GLPASELEFNLFN 148

Db 97 DPFAPETVVVTCDVIEPSTGQYERDRPSIARAEYKLGCTIGTAFGPEPEFVFD 156

QY 149 ET-----YKSASQKHWKNLK-----TAQPHQWNN 173

Db 157 EVKWDIDMSGARHTLIAEAAWSTGKDYSGNSGHRPRVKGGYFPVPVDSAQD---MR 212

QY 174 ISASSGIEFTMRVRKNLEAGILMEATHEPFLPSQHELNFPADPLTWADRHIIAKGV 233

Db 213 AEMCAKIEDIMGPR-----VEVHHHEVASQLEIGVSFTLVLRKADEVQQPKYAV 263

QY 234 REMAEGSGVATPMAXLSTALGNACHIMSLQDAETEKNAFYDQNDVEY-GMSTLARNMI 292

Db 264 WNVHAYQVATPMKPMVMDNGSGMHVMSI---SKDGKPLF--AGDEVAGISEMALYFI 319

QY 293 AGLLKYVPEATYFFASYINSYKRLQPLTFAPTCKCWAIDNRITSAPRL-CNSKSEGINVEL 351

Db 320 GGIHKHARALNATNPSTNSYKELVPHFEAPIMLAYSARNRSASIRIPVSNPKGKRIA 379

QY 352 RIGGADLNPLAPSAIIAGISGIEKLELPPPPASGNVN-----DKELPEFPNSLQAT 406

Db 380 RFPDPMNPLGPAALLMAGIDGIONKIHGPEAADKNLYDLPPPEEAKIPTVAHSLDMAL 439

QY 407 HLLK---ESKMLNKTGKILHYVNAANVEINEFS 439

Db 440 EALQADHFLKGGVTKEMLDAYIELKTEDVRRLN 475

RESULT 13

US-09-107-532A-4613

; Sequence 4613, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESS: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660

; COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4613:

SEQUENCE CHARACTERISTICS:

LENGTH: 190 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1...190

SEQUENCE DESCRIPTION: SEQ ID NO: 4613:

US-09-107-532A-4613

Query Match 10.2%; Score 241.5; DB 4; Length 190;

Best Local Similarity 29.5%; Pred. No. 2.1e-18;

Matches 56; Conservative 38; Mismatches 89; Indels 7; Gaps 2;

QY 272 KNAFYDQNDVEYGMSTLARNWIAAGLLKYVPEATYFFASYINSYKRLQPLTFAPTCKCWAID 331

Db 1 KKCLYDKGELESETAYHFLGGLLKARAYTAVCNPTVNSYKRLVPGVEAPVYVWMSGR 60

QY 332 NRTSAFRCLNSKSEGINVELRIGGADLNPLAPSAIIAAGISGIEKLELPPPPASGNVY- 390

Db 61 NRSPLVRVPSRGLSTRLESLRSDPSANPYLAMAVLQAGLDGIRNQITPPPAVDNRNIYV 120

QY 391 NDKE-----LPEFPNSLQATHLKESKMLNKTGKILHYVNAANVEINEFSKQVTD 444

Db 121 MDEEREAHIOQLPSTIHNAIKELRKONVMIDALGQHIFSNFVBAKLEWAAFRQTVSE 180

QY 445 WELNQGFNRY 454

Db 181 WEREQYLELY 190

RESULT 14

US-09-107-532A-6200

; Sequence 6200, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESS: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660

; COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107.532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6200:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...275
SEQUENCE DESCRIPTION: SEQ ID NO: 6200:
US-09-107-532A-6200

Query Match 5.6%; Score 134; DB 4; Length 275;
Best Local Similarity 23.0%; Pred. No. 3.6e-06;
Matches 63; Conservative 33; Mismatches 108; Indels 70; Gaps 9;

QY 3 ITYDELNNLRNGKIDTVLVACVDMQRLMGKRLTGRHFLGLDQKKISITFYVATIEG 62
DB 12 ITGEEIKRIIEENVRFLRMFTDILGTIKNVVPSQIDKVLKXMPD-----GSSIEG 67

QY 63 IAGGVEISSVDVGYSDCHLCADLNSLHLLPWSEG-----AVLAISNPHNFTVTSPLFC 116
DB 68 F-----VRIESDMWLYLPDLSTWMPFWSAAGKVARLICDIYNP-----DGTFFAG 114

QY 117 SPRVILMQQIERLANLK--LKGLFASLEFNLFNFTYKSASQKHWNKLTAKTAOPHHQWMI 174
DB 115 DPR-----GNLKRALKDM--RDLGFTSN-----LGPEPEFFFLKL 148

QY 175 SASSGIETFM-----RSVRNKLEBEAGILMEATHPEPLPSQHEINLV 215
DB 149 DEDGGITTLNDKGGYDFAPTDLGNCRRDIVLESLGLFEVEASHHEVAPGQHEIDFK 208

QY 216 PADPLTMADRHIIAKHGVREMAEQSGMVATFMAK 249
DB 209 YADVVDACDNIQTFFKLIVKTIARKHGLHATFMFK 242

RESULT 15
US-09-328-352-4580
Sequence 4580, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4580
LENGTH: 481
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4580

Query Match 4.0%; Score 95; DB 4; Length 481;
Best Local Similarity 20.0%; Pred. No. 0.21;
Matches 84; Conservative 69; Mismatches 170; Indels 98; Gaps 17;

QY 82 LCADINSLLHLPWSECAVLAINSPHNPVTSEPLFCSPRVILMQQIERLANLKLG--LFA 139
DB 81 LCANLNLIFKKEMFEAHIIHELEALMNGKTH---FSSTTINYTLGKRI-DVQLRGAILPG 136

QY 140 SELEFLNLET-----YKSASOKHWKLNKTAQ-----PHQWMNISASSGIEETFRSV 187
DB 137 SETTFRILITTEDITPYQNALRQBEKNRRLAESMFIYSFTSLM-----VEDFSR-I 187

QY 188 RNKLEBAGIL-----MEATHPEPLPSQHE-----LNFVPADPLTMADRHIIAKHGVR 234
DB 188 KNRIDQLRLGIEDFRFLDVHPEFVRQCIEDILIDVNVQSTLDLKFAPDKTLLKNTHK 247

QY 235 EMAREQSGVATPMAKLSSTALGNACH-----IHMSLODAETEKNAFYDQNDVEYGM-- 284
DB 248 IFABE--MVETFREQLIELWQGNIIHKKREAVNVALDQGIIRNVLLQTFVFPGEETWGLVQ 305

QY 285 ----STLARNWITAGLLKYYPE-----ATYFFASYINSYKRLQPLTFAPTCKCWAIDN--R 333
DB 306 VALTDITARKKAENYLEYLGQHDVLTCLYNRAFFTTELRLNRSIIRPVSSIFLDMNGLK 365

QY 334 TSAFRLCNSKSEGINVELRIGGADLNPYLAFSAIIAAGISGIEEKELELPPPSAGNVY--- 390
DB 366 ETNDQLGHDIGDGL---LRRVGNILNQAIMNTAYTASRIGGDEFVILMPGADANVLMML 422

QY 391 -----NDKELPEFPNSLQNAATHLLKESKMLNKTPEKLI 424
DB 423 QTIQELFNIDNQYSSHPISIAIGHATTKEQVEDMLKRAHHMYQKK---KSYTQEIL 479

QY 425 L 425
DB 480 L 480

Search completed: December 17, 2003, 22:45:22
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:37:15 ; Search time 39 Seconds
(without alignments)
1847.740 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDLNLLIRNGKIDTV.....INBFKQVTDWELNQGFNRY 454

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A Geneseq_19Jun03.*
- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
 - 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
 - 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
 - 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
 - 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
 - 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
 - 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
 - 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
 - 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
 - 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
 - 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
 - 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
 - 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
 - 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
 - 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
 - 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
 - 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
 - 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
 - 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
 - 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
 - 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
 - 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
 - 23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
 - 24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	614.5	25.9	459	22 AAG80030	Pseudomonas sp KIE
2	614.5	25.9	459	22 AAG80038	Pseudomonas putida
3	502.5	21.1	472	22 AAU54831	Propionibacterium
4	475.5	20.0	446	22 ABP33500	Brevibacterium lac
5	470.5	19.8	446	22 AAG92188	C Glutamicum prote
6	470.5	19.8	446	22 AAB79681	Corynebacterium gl
7	464	19.5	446	23 ABB55555	Lactococcus lactis
8	453	19.1	445	23 ABP66039	Bifidobacterium lo
9	445	18.7	448	23 ABP27862	Streptococcus poly

10	441	18.6	448	23	ABP27863	Streptococcus poly
11	432	18.2	448	22	AAU37785	Streptococcus pneu
12	432	18.2	448	24	AAE33506	Streptococcus pneu
13	432	18.2	448	24	ABU00865	S. pneumoniae type
14	424.5	17.9	446	22	AAU34237	Staphylococcus aur
15	424.5	17.9	446	22	AAU36745	Staphylococcus aur
16	424.5	17.9	446	22	AAU37182	Staphylococcus aur
17	410	17.3	444	23	ABB48084	Listeria monocytog
18	409	17.2	446	22	AGB82578	S. epidermidis ope
19	409	17.2	446	22	ABP39328	Staphylococcus epi
20	402	16.9	451	22	AAU34921	Enterococcus faeca
21	400.5	16.9	443	10	AAF93191	Peptide with glut
22	393.5	16.6	446	19	AAW37905	Mycobacterium tube
23	389.5	16.4	483	22	ABP96475	Putative P. abyss
24	333	14.0	845	23	ABB92646	Herbicidally activ
25	322	13.6	477	22	AGB93231	C Glutamicum prote
26	322	13.6	477	22	AAW79682	Corynebacterium gl
27	306	12.9	478	18	AAW18160	Mycobacterium tube
28	305.5	12.9	478	23	ABP65893	Bifidobacterium lo
29	303	12.8	478	19	AAW63035	Mycobacterium tube
30	303	12.8	478	19	AAW37901	Mycobacterium tube
31	294.5	12.4	533	22	AAU54574	Propionibacterium
32	287	12.1	472	22	AAU35544	Haemophilus influe
33	287	12.1	1038	22	ABG24868	Novel human diagno
34	281.5	11.8	500	24	ABP57636	S. murayamaensis A
35	265.5	11.2	469	22	AAU36495	Pseudomonas aerugi
36	265	11.2	481	22	AAU35914	Helicobacter pylori
37	255.5	10.8	524	24	ABP80864	N. gonorrhoeae ami
38	247	10.4	469	22	AAU38196	Salmonella typhi c
39	241	10.1	469	22	AAU34825	E. coli cellular p
40	216.5	9.1	816	22	ABG13707	Novel human diagno
41	201.5	8.5	476	22	ABG25229	Novel human diagno
42	180.5	7.6	185	19	AAW85879	S. pneumoniae deri
43	169.5	7.1	334	22	AAU23182	S. pneumoniae deri
44	160.5	6.8	118	23	ABP32638	Novel human enzyme
45	147.5	6.2	159	21	AAW41477	Human ORF1611 prot
						Human ORFX ORF1241

ALIGNMENTS

RESULT 1

AAAG80030
ID AAG80030 standard; Protein; 459 AA.
XX
AC AAG80030;
XX
DT 17-JAN-2002 (first entry)
XX
DB Pseudomonas sp KIE171-BII ipuC protein.
XX
KW L-alaninol; isopropylamine; iup1; ipuH; ofloxacin; ipuC; glutamylamide;
KW gamma-glutamylamide synthase; theanine.
XX
OS Pseudomonas sp.
XX
PN WO200173038-A2.
XX
PD 04-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-EF03651.
XX
PR 31-MAR-2000; 2000EP-0106888.
XX
PA (LONZ) LONZA AG.
XX
PI Leisinger T, Van Der Ploeg J, Kiener AM, De Azevedo Waesch SI;
PI Maire T;
XX
DR WPI; 2001-626266/72.
XX
DR N-PSDB; AAI68686.
XX
PT New microorganism for converting isopropylamine to L-alaninol, useful

QY 65 GGGVEISSVD---TGYSDDLCHLADLNSHLHLLPMSG-----AVLAISNPHNFVTSPL 114
 D 114
 D 48 GLGFDGSAIEGWTVRSEDDMIVQDPSTFQILPWRGGPGQTARMFCDLILTP-----DGEPS 103
 QY 115 FCSRVILMQOIERLANLKLGL---FASLEFNLFNFTYKASQKHKNLKTAPPHQW 171
 D 171
 D 104 LGDRHVLKR---TLAKAKDKGFTFYVHPHEIFLFE-----SQDDWSKAPTIDEQGY 154
 QY 172 M-NISASGIEFTWRSVKNKLEAEAGILMEATHPEFLPSQHELNFVPADPLTMADRHITAK 230
 D 230
 D 155 FDHVPSPGMD-FRATVNMLEQWISVEYSHHAGFCQNDILRYADALTMADNIMTFR 213
 QY 231 HGVREMAQSGVATFMAKLSSTALGNACHIMSLQDAETKNAFYQDNDYGHMSTLARN 290
 D 290
 D 214 TWKEISLERGIHASFPEKPLADAPGSGMTHLSL--PEGDSNAFYEAQGFNNLSLTARQ 271
 QY 291 WIAGLLKVPFATYFFASYNSYKRLQPLTAPTKCCWAIDNRTSAFRLCN---SKSEGI 347
 D 347
 D 272 FAAGILYHAAEICAVTDQVNSYKRLMGNEAPSYICWGHNNRSALLRIPOYKPKGNLSA 331
 QY 348 NVELRIGGADLNPYLAFSAIIAAGISGIEEKLPLPPASGNVY--NDKE-----LPEPPN 400
 D 400
 D 332 RMEFRALDPVANPYLAYSVLAAAGLDGDKQWQGEPTSDSDVWELTDCGERQAMGIQPLPR 391
 QY 401 SLQNAHLHLLKESKMLNKTGKELILHYVNAANVEINEFSPKQVTDWELNQ 449
 D 449
 D 392 SLDEALKIMKESDFADVGLGSHAFGYFLDNKHQEWEEYNQQVTPVELKK 440

RESULT 9

ABP27862
 ID ABP27862 standard; Protein; 448 AA.

AC ABP27862;
 XX

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 4900.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABN68493.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 1; Page 3652; 4525pp; English;

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX Sequence 448 AA;

QY Query Match 19.7%; Score 445; DB 23; Length 448;
 D Best Local Similarity 27.4%; Pred. No. 2e-34;
 D Matches 131; Conservative 80; Mismatches 213; Indels 54; Gaps 10;

QY 1 MTITYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQ---KKISISTFYVA 57
 D 57

QY 1 MTITAEIRREVKEKNVTFRLMFTDILGVNKNVEIP-----ATDEQLDKVLSNKAMPDG 55
 D 55

QY 58 VTIEGIAGGGYEISSVDYGYSDCHLCADLNSHLHLLPMSE--GAVLAISNPHNFVTSPLF 115
 D 115

QY 56 SSIEGF-----VRINESDMVLYPDLOTWTFVPMGDENGAVAGLICDIYTAEGEPFA 106
 D 106

QY 116 CSRVILMQOIERLANLKLGL-FASELEFNLFNFTYKASQKHKNLKTAPPHQWNNI 174
 D 174

QY 107 GDRGNLKRNMKRMQEMGYKSNLGPPEPFLF-----KMDENGCPNPTLDV 151
 D 151

QY 175 SASG-----IETFMRSVRNKLSEAGILMEATHPEFLPSQHELNFVPADPLTMADR 225
 D 225

QY 152 NDKGGVFDLAPTLADNTRREIVNLTQMGEFVASHHVAVGQHEIDFYDDVYLKACDN 211
 D 211

QY 226 HIITAKHGVREMAQSGMVATFMAKLSSTALGNACHIMSLQDAETKNAFYQDNDYGM- 284
 D 284

QY 212 IQFLVVKVTKIARKHGLYATFMAKPKPGINGSGHCHNMSLFDNE-GNNAFDPEDPRGMQ 270
 D 270

QY 285 -STLARWIAGLKLYPEATYFFASYNSYKRLQPLTFAPTKCCWAIDNRTSAFRLCNSK 343
 D 343

QY 271 LSEDAVYVFLGGLMKHAYNTAIINPTVNSYKRLVPGVEAPVYVAVAGNRSPILRVPASR 330
 D 330

QY 344 SEGIVNELRIGGADLNPYLAFSAIIAAGISGIEEKLPLPPASGNVY-----NDKELP 396
 D 396

QY 331 GMGTRLELRSDVPTANPYLALSVLGSGLEGIEKIEAPEPIETNIYAMTVEERRQAGIV 390
 D 390

QY 397 EFPNSLQNAHLHLLKESKMLNKTGKELILHYVNAANVEINEFSPKQVTDWELNQGFNRY 454
 D 454

QY 391 DLPSTLHNALEAEDEEVVKAALGTHIYTNFLQAKRIEWASYATYVSQWEIDNYLDLY 448
 D 448

RESULT 10

ABP27863

ID ABP27863 standard; Protein; 448 AA.

XX ABP27863;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 4902.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus pyogenes.

XX WO200234771-A2.


```

PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-JB02163.
XX
XX 27-MAR-2001; 2001GB-0007658.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX N-PSDB; ABX06145.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to Streptococcus bacteria, such as pneumonia, sepsis, otitis media
XX or ear infection
XX
XX Claim 1; SEQ ID No 866; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions' from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX ABS56454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2469 proteins expressed by the identified coding regions from the
XX genomic sequence.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIFO at
XX ftp.wifo.int/pub/published_pct_sequences.
XX
XX Sequence 448 AA;
XX
XX Query Match 18.2%; Score 432; DB 24; Length 448;
XX Best Local Similarity 27.2%; Pred. No. 3.7e-33;
XX Matches 128; Conservative 83; Mismatches 220; Indels 40; Gaps 10;
XX
XX 1 MTITYLNNLRNGKIDTVLACVDMQRLMKRLTGRHFLGLDQ---KKISISTFVYA 57
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 1 MPITAADIRREVKEKNVFIIRLMPDSILGTMKNVIP-----ATDEQLDKVLNKMVFDG 55
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 58 VTIEGIAGGVEISSVDTGYSCHLCADLNSLHLLPWSE--GAVLAINPNHNVFTSBLFL 115
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 56 SSIEGF-----VRINESDMYLPDLDTVTFPFGDENGSVAGLICDVVTEGEPPA 106
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 116 CSPRVLVQQIERLANLKLGL-FASELEFNLF--NETYKSASQKHWNKLTQAQHHQM 172
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 107 GDRGNLKRALRHMBEVGPKSNLGEPEFFFLKLDENGDPFLVNDKG-----GYF 158
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 173 NISASGIGTFRWSVRNKEEAGILMEATHPEFLPSQHELFNVPADPLTMADRHIIAKHG 232
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

159 DLAPTDLADNRREIVNVLTQMGPEVEASHHEVAVGQHEIDFKYDEVLRACDKIQIFKLV 218

233 VRENAESGMVATFNAKLSSTALGNACHIHMSLODAETEKNAFYDQNDVEYGM--STLARN 290

219 VKTIARKHGLYATFMAKPKFGIAGSGMHCNNSLFDAE--GNNAFFDPNDPKGMQJSETAYH 277

291 WIAGLLKYVPEATYFFASYINSYKRLQPLTFAPTKCCWAIIDNRTSAPRLCNKSEGINVE 350

278 FLGLLIKHANYTAIMNPPTVNSYKRLVPGVEAPYVIAWAGNRSPLRVVPASRGMGTLE 337

351 LRIGGADLNPLYAFSAIIAAGISGIEKLELPPPSAGNVY-----NDKELPFPFNSLQ 403

338 LRSVDPMPANPVVAMVLEVLGVLGKIEAPAPIEENIYIMTAERKEAGITDLPSTLH 397

404 NATHLLKESKMLNTFGKELILHYVNAANVEINFEFSQVTDWELNQGFNRY 454

398 NALKALTEDEVKAAALGDHIYTSFLEAKRIEWSAYATFVSQWEIDNLDLY 448

RESULT 14

AAU34237

ID AAU34237 standard; Protein; 446 AA.

XX

AC AAU34237;

XX

DT 14-FEB-2002 (first entry)

XX

DE Staphylococcus aureus cellular proliferation protein #513.

XX

KW Antisense; prokaryotic cellular proliferation protein;

XX

OS Staphylococcus aureus.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

XX 21-MAR-2001; 2001WO-US09180.

XX

PR 21-MAR-2000; 2000US-191078P.

XX

PR 23-MAY-2000; 2000US-206848P.

XX

PR 26-MAY-2000; 2000US-207727P.

XX

PR 23-OCT-2000; 2000US-242578P.

XX

PR 27-NOV-2000; 2000US-253625P.

XX

PR 22-DEC-2000; 2000US-257931P.

XX

PR 16-FEB-2001; 2001US-269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

XX Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

XX

DR N-PSDB; AAS52096.

XX

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX

XX Example 3; Seq ID No 5733; 511pp; English.

XX

PS The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

Search completed: December 17, 2003, 22:42:57
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 17, 2003, 22:51:30 ; Search time 5280 Seconds
(without alignments)
3517.608 Million cell updates/sec

Title: US-10-098-602A-2

Perfect score: 2376

Sequence: 1 MTITYDELNLNRKIDTV.....INEFSKQVTDWELNQGFNRY 454

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10098602/runat.17122003.151235.28042/app_query.fasta_1.647
-DB=GenEmbl -QWTF=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10098602.cgn 1.1 3508 @runat.17122003.151235.28042 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgtg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	1015	42.7	345783	1	AP003001 Mesorhizo
C 2	969.5	40.8	11725	1	AE014569 Brucella
C 3	963	40.5	323450	1	SMES91790 Sincorhizo
C 4	949.5	40.0	12340	1	AE009691 Brucella
C 5	942.5	39.7	10069	1	AE009164 Agrobacte
C 6	942.5	39.7	11253	1	AE008130 Agrobacte
C 7	799.5	33.6	13246	1	MTV003
C 8	799.5	33.6	14715	1	AE007117 Mycobacte
C 9	799.5	33.6	318050	1	BX248344 Mycobacte
C 10	798.5	33.6	277000	1	SC093109 Streptomy
C 11	773.5	32.6	298550	1	AP005047 Streptomy
C 12	614.5	25.9	1860	6	AX259711 Sequence
C 13	614.5	25.9	11355	1	PSP311159 Pseudomon
C 14	595	25.0	128136	1	AF440524 Pseudomon
C 15	554	23.3	300550	1	AP005021 Streptomy
C 16	510.5	21.5	305153	1	AP001520 Bacillus
C 17	510	21.5	310266	1	AE016862 Pseudomon
C 18	509	21.4	300242	1	AE016790 Pseudomon
C 19	493.5	20.8	1338	6	BD180047 Highly th
C 20	487	20.5	10453	1	AE004758 Pseudomon
C 21	481	20.2	18194	1	AE001758 Thermotog
C 22	477	20.1	1444	1	X60160 T.maritima
C 23	475.5	20.0	5500	6	AX503498 Sequence
C 24	475.5	20.0	5500	6	BD177689 Process f
C 25	470.5	19.8	1338	6	AX122526 Sequence
C 26	470.5	19.8	1338	6	BD164643 Novel pol
C 27	470.5	19.8	1461	6	AX063813 Sequence
C 28	470.5	19.8	6433	1	CGL310086 Coryneb
C 29	470.5	19.8	306650	1	AP005221 Coryneb
C 30	470.5	19.8	326150	1	AP005281 Coryneb
C 31	470.5	19.8	349980	6	AX127150 Sequence
C 32	466	19.6	2192	1	CLOGINSA
C 33	466	19.6	142665	1	AE016794 Pseudomon
C 34	465	19.6	10458	1	AE009688 Brucella
C 35	465	19.6	13418	1	AE014571 Brucella
C 36	465	19.6	310613	1	AE016875 Pseudomon
C 37	464.5	19.5	11906	1	AE001904 Deinococc
C 38	464	19.5	12185	1	AE006449 Lactococc
C 39	463	19.5	43883	2	AC004396 Pseudomon
C 40	459	19.3	12795	1	AE004467 Pseudomon
C 41	458	19.3	9960	1	AE004630 Pseudomon
C 42	453	19.1	10876	1	AE014758 Bifidobac
C 43	453	19.1	11364	1	AE012342 Xanthomon
C 44	453	19.1	305245	1	AE016793 Pseudomon
C 45	453	19.1	349980	6	AX492786 Sequence

AP003001 indicates GS

ALIGNMENTS

RESULT 1

```

AP003001/c
LOCUS       AP003001             345783 bp    DNA    linear    BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 8/21.
ACCESSION   AP003001
VERSION     AP003001.2  GI:14023393
KEYWORDS    .
SOURCE      Mesorhizobium loti
ORGANISM    Mesorhizobium loti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1 (bases 1 to 345783)
AUTHORS     Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Iidesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
            Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
            DNA Res. 7 (6), 331-338 (2000)
MEDLINE     21082930
PUBMED      11214968
REFERENCE   2 (bases 1 to 345783)
AUTHORS     Kaneko,T.
TITLE       Direct Submission
JOURNAL     Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
            On May 11, 2001 this sequence version replaced gi:11994976.
FEATURES             Location/Qualifiers
     source           1..345783
                     /organism="Mesorhizobium loti"
                     /mol_type="genomic DNA"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
                     complement(6..884)
                     /gene="mlr3011"
                     complement(6..884)
                     /gene="mlr3011"
                     /codon_start=1
                     /transl_table=11
                     /product="transcription regulator"
                     /protein_id="BAB50000.1"
                     /db_xref="GI:14023394"
     gene             /translation="MDRLMAMSTFIRVIETGFSAAARHLRVQGSXKVAQLEERL
                     GRLLSRTTGLSPTDAGLSFYEGAKRAVDDEADEAENAAAGTAALTLGLRIGTSVTF
                     GRLLVPRKLFADHPRLDVEVAMDRNVDLVEGVDMRLRLGESFAPAHIGRKIAA
                     CRLLVVAAPSLAREGEPIAPDELSRHEAILFVRAGTSIRWTFRRGAEVSLAGGRL
                     RYTSAGIRENLVSGIGFTVGSSEWFDALRSRGAVRPILTEWTLPPVELWALFPAGRK
                     ASARATVFELAGSL"
     CDS              994..1740
                     /gene="mlr3013"
                     994..1740
                     /gene="mlr3013"
                     /codon_start=1
                     /transl_table=11
                     /product="short chain dehydrogenase or 3-ketoacyl-CoA
                     reductase"
     gene             /protein_id="BAB50001.1"
                     /db_xref="GI:14023395"
     CDS              /translation="MGKLNKNAIVTGAAIGTAGIATAKLAEGAAGVNVNATSRBA
                     ERTKSVIAGGKATAVQGVDAEAEVEKFAAEIAPKPSILINNAATASFGTFEE
                     ASEADYRFLDFNLVTLTITKAAMRHPFTETGSIIVNIGTISWNPNPDIYSASKA
                     AIDLTLSLRGARNIRVNIAPGYTHETMETEGMVGTFDGNLTIAGVPLGQRFQK
                     DDIAPTVAFLADEAAWLTERINAGSGAR"
                     1914..3371
                     /gene="mlr3014"
                     1914..3371
                     /gene="mlr3014"
                     /codon_start=1
                     /transl_table=11
     gene             /product="homospermidine synthase"
                     /protein_id="BAB50002.1"
                     /db_xref="GI:14023396"
                     /translation="MLDKMANENHPVYGEITGPVVMIGFSGISGRTLTPLIERHFKFDK
                     SRMTVLDPDRDLDRKLLDGERIAPVQEAFTENNYKLLRPLLTNGGGGQFCVNLSDVTG
                     SVDMRLCRKLGVLVYIDTVBFWLGFYEDAKADNARNRTNYALREAMIKEKDKPGGAT
                     VSTCGANPGMVSWFKQALVNLATDLGFESEPAQEDREGAKMLKMGKVKGHGIAE
                     RDTORTKKPKMDVFWNTWSVEGFTISEGLQPAELGWGTHEKMMPKNGKGHKHSKAAI
                     YLEOPGANTRVRSNCPGACQYGLLVTHNEAISIAIDFTVRSKKGKGVYRPTCHVAYH
                     PSNDANLSLDEMFGAAGKAPQVHHVLDENELVDGDELGVLLYGHDKNAYWGSLSL
                     ABARKLAIQNAQTAGQVTSVLAGVMVWALENPDAGIVEADEMDYKRCILDVQSPYLGVP
                     KGYTDMTPFLDRRPLGFPEDLDRSDPPWQFRNILVR"
                     3431..3931
                     /gene="mlr3016"
                     3431..3931
                     /gene="mlr3016"
                     /codon_start=1
                     /transl_table=11
                     /product="outer membrane lipoprotein"
                     /protein_id="BAB50003.1"
                     /db_xref="GI:14023397"
                     /translation="MRPGAACVLDRDARPFQHFGRSHRACQNNRFLDRCVWRGFMQFTI
                     ARKVLISGMAGSFAMLAACITTSQDPDAPMAAPKGVGSGWIDAKGTGLSTPTAGKFA
                     TVATDTGQKLADGSYTMGTGATSVETNGTSLIRQTPVFNCLLISTSOLNCTSSSQGNF
                     VLTRRT"
                     4123..6243
                     /gene="mlr3017"
                     4123..6243
                     /gene="mlr3017"
                     /EC_number="3.1.3.5"
                     /codon_start=1
                     /transl_table=11
                     /product="s'-nucleotidase"
                     /protein_id="BAB50004.1"
                     /db_xref="GI:14023398"
                     /translation="MKKLAALAAALSVTLGLSAGASPADYTLNLIHFNMDHMSIRIGNN
                     KYESTCSADEETKGEICIGAGRLITAIQAERKKLEQNVLLNAGDSFGSLFYTTYK
                     GTVEEFLNOMKPDVTLGNHEFDDESALVPYLDKAKFPPIVSANMPPDKSAGAKKI
                     KPSIVVEVGOKIGIVGAVTNDTPELASPGNIIAADDVESITAEVEKLKAQGVKVI
                     AVTHIGYNRRERDIIAKIPGVVVVGGHSHILLSNDTPKAAGPYPTMVDNPGYKVPV
                     QAASYKYLGEFVVDNDNGVVSASGDPFLDKRSITDPDPVAVLARIKELGALIEALKN
                     KEVATETVIDGSRNCRACECMGNLVSDAILDRVQGVGEIVISNGGGLRASIDKG
                     TVTMGEVLTVLPQNTLATFQISKDLVAGLESLSQIEDGAGRFQVAGLKYSPDKS
                     VAPNAGRVKSVSEVMENGAWTPIKPKDYLVTATNNYVROGGDGYKYPAAKNAKYDGP
                     GLEOVVADYLGARHPYTKLDGRITETIAATVAAAAPAEPAKPAEATPAPAPAKPA
                     EAKAPAEPAKPAEAPAPPELPANGNIANTTPAISTEAPPPAATPAPAEPAK
                     AEAPAPAEPAVPVPAEPAKPAEAPAEPAATGSHVIVAGDTITWDLAKKAYGATKWK
                     IYEANKGQRPHRLTPGATLTTPAK"
                     6381..7442
                     /gene="mlr3019"
                     6381..7442
                     /gene="mlr3019"
                     /EC_number="4.99.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="ferrochelatase"
                     /protein_id="BAB50005.1"
                     /db_xref="GI:14023399"
                     /translation="MTSAIADPKAAKPVGTAGHAPAGAGKGVLLVNLGTPDGTGF
                     KPMRYREFLSDPRVIELNKAIWYPILYGLVLTTRPKSGANYARIWNQERNESPLR
                     TVTRAQSEKLAEALDGLDPTVTDWAMRYGNPSTASVARELVAQGDRIILSFPLYQYS
                     ITTATANDQLFRALMKRRRAPRIASVPVYAEVPIEALASSIERHLATLDFEPEV
                     ATTVHGIPKPYSDKGDQPYQACHLETTRLLREKLGWDEKLIITTFOSRGAQEWLQPYT
                     DKTVSKLGDGVKSAIYNPFSVDCITELDEIGREAAETFHAGSKGPAHIFCLNDS
                     AEGMTVIAAMVRRRLSGWV"
                     7615..8826
                     /gene="mlr3020"
                     7615..8826
                     /gene="mlr3020"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB50006.1"

```

```
/db_xref="GI:14023400"
/translation="MPKRLDLRGRPWMSAYRAPAVTSALTREDVTTDLVLIQMGISG
AMAEALATDSVHSVCDIRPRGPKGSTAATLAVQFEIDQPLSLSKMGAKAQAW
RRRLATSLNGLIEDLAICRLRRTPSLYLAGTLGPASMRDEAEARQAGAAATVL
TPOLAQTGIDRGATLSHGNALDPRKLTAGLLKALERKARLAPAEATLEDA
GEAVATKDGPTLTARHVLTATGVELDVIPAAHRLIISTWAIATCPQPKLWPGAF
IWEASDYLRLATADRVIICGGEDEFADETRDELIAKDSARIADKGLRFLPYLDV
RPFATWGSFGTTSAGLPIYIGAVPGHPRIHAVMGVNGNGITFSRIASEIISASIKGSA
DTDAELPAFNR"
gene      8948..9898
CDS       /gene="mlr3021"
          /note="hypothetical protein"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB50007.1"
          /db_xref="GI:14023401"
          /translation="MDFSGFDIAIGALALLVLVLLIKGIRTIPOGVNVTVERGRVTK
          TSPGNFIFFVDRIKAGNMMEQVLDPSEIITRDNAIVGVGDGIAFFQILNAAQA
          AYQVGLQNALILMTNRTIVMSGMDLDELNLDAINERLLRVVDDEAAHPWGIXIT
          RVBIKQINPPANLIESGROWTAERNKRAQILAAEGLKQSOILEAEGRKEAARDAEA
          RESAABAEARATVSEAIKSGDVOALNYFAOKYTEALGKIGTATNSKIVLMPFEAS
          SLJSGLGIGIEIAKEVFRSEGTGAQRQAARPPVVRPEN"
          9912..10376
          /gene="mlr3022"
          9912..10376
          /gene="mlr3022"
          /note="hypothetical protein"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB50008.1"
          /db_xref="GI:14023402"
          /translation="MFDRIISLPGNMNMLVFLVLMVEIIAPGIFMLWIGIAALIIG
          AVSLIWDAAITWQVQLAFLALSLVAVVKGKLMGGRDPTDPLNPRGQWVR
          MATLAPKDGRIKGLDTLWRVSGPDLPAQTQVRVTSADTDLELTVRV"
          complement(10629..11630)
          /gene="mlr3023"
          complement(10629..11630)
          /gene="mlr3023"
          /note="hypothetical protein"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB50009.1"
          /db_xref="GI:14023403"
          /translation="MARSUDKKPLDRQASIDSALRTVATEQAGIAALAPALENGLAA
          PFAQVDMISKGRILVITGVKSGHIGSKIAATLASTGTPPAFFVHPVVEAHGDLGMI
          AKDIAIATSWGSEKMLGIVAYSRFSIPLIATVSGETSALARADAVLLLPRTPE
          ACPHGLAPTITLLQVLIGDALAILEARGFTPDHFRTPHGGQLGANLTVWSEIMR
          VGDQMLAVLGTMPKPAVMTLSQKVGCVLIVDANGELAGIITDGDVARNLHRNLADV
          IVDVMTPTKTPDTPQTLAGTATALLNEHNIGALVTRNNMPLGVVHFHDLRIGAA"
          11830..13626
          /gene="mlr3024"
          11830..13626
          /gene="mlr3024"
```

Alignment Scores:

Pred. No.:	3,29e-82	Length:	345783
Score:	1015.00	Matches:	206
Percent Similarity:	63.64%	Conservative:	81
Best Local Similarity:	45.68%	Mismatches:	164
Query Match:	42.72%	Indels:	0
DB:	1	Gaps:	0

US-10-098-602a-2 (1-454) x AP003001 (1-345783)

```
Qy      4 ThrTyrAspGluLeuAsnAsnLeuLeuIleArgAsnGlyLysIleAspThrValValLeuAla 23
          ::::::::::::::::::::
Db      61983 TCGTTCGATCAGTTGAAGAAAGCGGTCTCCAGCGGTGAGATCGACACGGTGTGGCGTCG 61924
          ::::::::::::::::::::
Qy      24 CysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGly 43
          ::::::::::::::::::::
Db      61923 ATCGTCGACATCGAGGCGCGGTCTCGCGGAAAGCGGTCTCGCGCAGTACTTCGTCGAT 61864
```

```
Qy      44 LeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGluGlyIle 63
          ::::::::::::::::::::
Db      61863 TCCGCGCATGACGAAACGATGCTCAACTATCTGCTGGCCCGCCGACATCGATATGGAG 61804
          ::::::::::::::::::::
Qy      64 AlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeuCys 83
          ::::::::::::::::::::
Db      61803 CCGGTCCCGGCTACAAAGCGCGGAGCTGGTGAAGAGGCTATCGCCACTTCGTCATGAAG 61744
          ::::::::::::::::::::
Qy      84 AlaAspLeuAsnSerLeuLeuLeuProTyrPheSerGluGlyAlaValLeuAlaIleSer 103
          ::::::::::::::::::::
Db      61743 CCGGACTGCGCCACGCTGCGGCGCATTCGTGGCTCGAAAAGACAGACACTTGTGATCTGC 61684
          ::::::::::::::::::::
Qy      104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123
          ::::::::::::::::::::
Db      61683 GACGTGCTCGATCACCACACCCATGACGACTCCGCGCATTCGCGCGCGCATCTCTGAAG 61624
          ::::::::::::::::::::
Qy      124 GlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGlu 143
          ::::::::::::::::::::
Db      61623 AAGCAGGTCAAGCGGCTCTCCGAGCGCGCTATATCGGCTATTTCGTTCCGAGCTGGAA 61564
          ::::::::::::::::::::
Qy      144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisIleTyrLysAsnLeuLys 163
          ::::::::::::::::::::
Db      61563 TTTACTCTGTTCAACGAGACCTATGATTCCGCCCGCAAGAAACACTGGCAGGCGCTCGAT 61504
          ::::::::::::::::::::
Qy      164 ThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPhe 183
          ::::::::::::::::::::
Db      61503 ACCGCTCGCCCTATATCGCGCATTCAGATCGGCATCACCACCAAGGAAGAGCGGTC 61444
          ::::::::::::::::::::
Qy      184 MetArgSerValArgAsnLysLeuGluAlaGlyIleLeuMetGluAlaThrHisPro 203
          ::::::::::::::::::::
Db      61443 ATCGTTCGGCTTCGCAACGAGATGGAAGCGCGCGCATCCCGATCGAGAATTCAGAGGC 61384
          ::::::::::::::::::::
Qy      204 GluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAla 223
          ::::::::::::::::::::
Db      61383 GAGTGGGCGCGCGCGCAGGAAGAGATCAATGTCGCTATTCGCGCGCTCGACATCGGCC 61324
          ::::::::::::::::::::
Qy      224 AspArgHisIleIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetVal 243
          ::::::::::::::::::::
Db      61323 GACCGCCACGTCATCTGAAGATGGCGCAAGGAGATCGCGAGTCCGAAGCAAGGCC 61264
          ::::::::::::::::::::
Qy      244 AlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMet 263
          ::::::::::::::::::::
Db      61263 ATCTCTTCATGTCCTCAAGTACAAATTACGGGCTCGCGCAATTCAGCCACATCCACAAT 61204
          ::::::::::::::::::::
Qy      264 SerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGly 283
          ::::::::::::::::::::
Db      61203 TCCTGTGAGCGCGCGCAGCGGACGCGCTTTGTTCTTCGACAAGAGCGGACTGGACG 61144
          ::::::::::::::::::::
Qy      284 MetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGluAlaThr 303
          ::::::::::::::::::::
Db      61143 CTGTCGACACTCGGCCCAAGATGGCGCGCGCGCGCTCAATATACCCCAAGGAATTCACC 61084
          ::::::::::::::::::::
Qy      304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
          ::::::::::::::::::::
Db      61083 TGGTTCCTGGCGCGCTTACATCAACTCCTACAAGCGCTTCAGCGCGCGCATTCGCGCCG 61024
          ::::::::::::::::::::
Qy      324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
          ::::::::::::::::::::
Db      61023 ACCAAGATCATGTGGAGCGGAGGACACCGCGCGCGGTTCGCTGTGCGCGCGAGGGC 60964
          ::::::::::::::::::::
Qy      344 SerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnProTyrLeuAla 363
          ::::::::::::::::::::
Db      60963 ACCAAGGGCATCCGATGGAATGCCGATCGCGCGCGCGCGCATCTCAATCCCTATCTCGCC 60904
          ::::::::::::::::::::
Qy      364 PheSerAlaIleIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGluLeuProPro 383
          ::::::::::::::::::::
Db      60903 TTCGCGCGCTGATCGAGCGCGCTCGCGCGCATCGATGAGAAGTCGAACTCGAACTCAAAA 60844
          ::::::::::::::::::::
Qy      384 ProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGln 403
          ::::::::::::::::::::
Db      60843 CCTTTCGTCGCGCGATCGCTATCAGCGCTCTCGCTCTCGCGCGAGATTCGGAAGAGCGTGC 60784
```



```
QY 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
Db 4942 GCGAAGGTACCAAGGCGATCGCATCGATCGCGATCGCGCGGGATATCAATCCC 5001
QY 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
Db 5002 TATCTGGCCTTTCGAGCGCTCATCGCGCGCGCTGAAAGCGCTGGATGAAAGCTGAA 5061
QY 381 LeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400
Db 5062 CTGACGAACTTTTGGCGCGATGCTATAGCGCGGTGAGGTGAGGAATTCCTTAC 5121
QY 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420
Db 5122 ACCTGCGCGAGCGCGCTCGCGGCTCAAGGTTCGCTTCTCGAAGAGCGCTTCGCG 5181
QY 421 GluLysLeuIleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLys 440
Db 5182 GAGGATGTCGTGAACATTACACCCATACCGCCCATTTGGGAGCAGATCGAATATGACCG 5241
QY 441 GlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArgTyr 454
Db 5242 CGCGTGACGACTGGGAGCTCTACCGCGGATTTGAACGCTAC 5283

RESULT 3
SMES91790/c
LOCUS SMES91790 323450 bp DNA linear BCT 05-JUL-2002
DEFINITION Sinorhizobium melliloti 1021 complete chromosome; segment 9/12.
ACCESSION AL591790 AL591688
VERSION AL591790.1 GI:15075230
KEYWORDS
SOURCE Sinorhizobium melliloti (Rhizobium melliloti)
ORGANISM Sinorhizobium melliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 323450)
Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boisard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portecelle,D., Puehler,A., Purnelle,S., Ramsberger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium melliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
21396507
11481430
Gouzy,J.
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/melliloti.html.
Location/Qualifiers
1 . 323450
/organism="Sinorhizobium melliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
155..1519
/gene="SMC01575"

FEATURES
source
155..1519
/gene="SMC01575"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC46787.1"
/db_xref="GI:15075231"
/db_xref="SPTREMBL:Q92K34"
/translation="WTPATRTSLSLFAALAMELAPANAQDAYGGYGGGDVMLVT
PGSDLDYIPGAEVHAMDRRTVLDVPMGNI VATVVPNDGQRRRREYGRDGY
PPPRDGYGSEPTGTAIPYRDIAPVEREDLPNSLSDREEAAYDPQDDP
LAQPPPMATVTGKRAEIALQVFLDREGFSGVIDGKMSNVTKAIEAWQATGET
LPNNATEDLERLRFNGGLPIITTYITTAADAAGPFVASIPDYAHKAQLPHLSFTSVT
EMLGKFMDEAYLRELNGVDFSPGTTIKVNPCKKGGKVARIVADKARKQVLAY
DRAGKLIAYPSTIGSSDTPSPGCTVHVERIAFDPGTYTNPKINFQGGANDRLQLQLP
GNGPVGVTWVIALSKPTYGTHGTPSPKIGTKQSHGCVRLTNWDATELGDMVSTGTVV
EFVD"
1707..2996
/gene="SMC01576"
1707..2996
/gene="SMC01576"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC46788.1"
/db_xref="GI:15075232"
/db_xref="SPTREMBL:Q92NJ3"
/translation="WATERPALNLMHATAPAPRTAPLAGDLTVEIVAIVGGGTGLSA
ALHAEFTGIRTAIVEMIGFGSGNRNGLVNGVMVQDDLIATIAAAGNRLLDEL
GGPSPFVLDVAKHGIECEAVRNGTLHMSVGAELKEIREREAQMKKCAPVEVLSAE
KAHLSGAGGFTGALLDRAGTIQPLAYARGIARALAAGAEIFTDPTLLAASRQDGL
WNLKTPRGCTVTHVILATNAYGSLVTGVPWKYRQELTILPYFQFATNPLDNPVAA
ILPERQAGWDTGLVMTSPMDRNLIFGSIQRDLDAIAAGTHRAFAARSVRKLFPYIG
DFAIFHWMDGRIGMTNNLPAHVLPANVVISGYNRGIAFGTVFGRALARHVTGDT
SAIPLAETPVPDPWRTLKSAFYHAGAQAHPIDKRF"
complement(3024..3740)
/gene="SMC01579"
complement(3024..3740)
/gene="SMC01579"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46789.1"
/db_xref="GI:15075233"
/db_xref="SPTREMBL:Q92NJ2"
/translation="MAVITFANAKGAGKTTAALILSTELARQGNRVVVLADAPQRWI
TWSVSGRVANLEVISHTVPASLPCHIRELGEADPIVIDLAGAKDAIVALAIGLSD
HVLIPVQGCAMDARGAIVOILEIRHIGEKARVINHSVLTNRVNSLVTTALQTIKAL
LASRGVSLDTPIVERVAREIFECGGTLQWMDPNRVSNLDKARENAYALAAAEVQNL
PVTARRALMSRUSALPRAA"
complement(3949..5094)
/gene="SMC01582"

Gene
```


Db 7258 ATCCACCAGTCGCTCTGGAGCAAGGACGGCGAAACCGCGCTGTCTTCGACAGAACGGG 7199

Qy 281 GluTyrGlyMetSerThrLeuAlaAraGlnTyrPheAlaGlyLeuLeuLysTyrValPro 300

Db 7198 CAGTACGGATGTCGAGCTCATGCGCCATATGTGCGCGGTGAGTCGCGCCATCGGAGC 7139

Qy 301 GluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr 320

Db 7138 GAGGTACCATTTCTCTGGCGCCCTATATCACTCTACAGGCGCTTCATGGCCGCTACC 7079

Qy 321 PheAlaProThrLysCysTyrPheAlaAlaAspAsnArgThrSerAlaPheArgLeuCys 340

Db 7078 TTGCTCCGACCAAGCGATCTGGAGCAAGGACCAACCGCCGCGCTATCGCTCTGCTG 7019

Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaPheLeuAsnPro 360

Db 7018 GCGAGGCGGACGAAGCAATCCGATCGAATCGCGCTGCGCGCTCCGATCTCAATCCG 6959

Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380

Db 6958 TATCTTGCCTTTCGCGCCCTGATCGCGCGCGGATTCGCGCATCGAGAACAGATGGA 6899

Qy 381 LeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400

Db 6898 CTGAGGCGCGTTCGTGCGGACGCTATCAGGCAAGGAGGTGCGCGAGATCCCGCAT 6839

Qy 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420

Db 6838 ACCGTGAGAGGCGGCGGAGCGCTCTCGCGCTCGAAGATGTCGCGCGCTTTCGCG 6779

Qy 421 GluLysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLys 440

Db 6778 GAAGAGTCTGATACCATATGTGCACGCGCAGATGGGACGAGCATATGACCG 6719

Qy 441 GlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453

Db 6718 CCGGTGACCATGCGGAGGTGCGCGCGGTTCGAAAGG 6680

RESULT 4

AE009691/c 12340 bp DNA linear BCT 20-MAR-2003

LOCUS Brucella melitensis 16M chromosome II, section 53 of 107 of the complete sequence.

ACCESSION AE009691 AE008918

VERSION AE009691.1 GI:17984724

KEYWORDS Brucella melitensis 16M

SOURCE Brucella melitensis 16M

ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella.

REFERENCE 1 (bases 1 to 12340)

AUTHORS DelVecchio,V.G., Kapral,V., Redkar,R.J., Patra,G., Patra,G., Mujer,C., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A., Reznik,G., Jablonaki,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S., O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyrpides,N. and Overbeek,R.

TITLE The genome sequence of the facultative intracellular pathogen Brucella melitensis

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)

PUBLISHED 11/5/6688

REFERENCE 2 (bases 1 to 12340)

AUTHORS DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA

REFERENCE 3 (bases 1 to 12340)

AUTHORS Elzer,P.H. and Hagius,S.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA

REFERENCE 4 (bases 1 to 12340)

AUTHORS Kapral,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,

Lykidis,A., Reznik,G., Jablonaki,L., Larsen,N., D'Souza,M., Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R., Kyrpides,N. and Overbeek,R.

Direct Submission

Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA

REFERENCE 5 (bases 1 to 12340)

AUTHORS Letesson,J.-J.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire, Laboratoire d'Immunologie et de Microbiologie, Universite of Namur, 61 rue de Bruxelles, Namur 5000, Belgium

REFERENCE 6 (bases 1 to 12340)

AUTHORS O'Callaghan,D.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Kennedy, Nimes 30900, France

FEATURES

source

1..12340

/organism="Brucella melitensis 16M"

/mol_type="genomic DNA"

/strain="16M"

/db_xref="taxon:224914"

/chromosome="II"

/complement(235..678)

/gene="BMEI10552"

/complement(235..678)

/gene="BMEI10552"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAL53794.1"

/db_xref="GI:17984725"

/translation="MNSFRKTCAGALILFGATSIPTVAAPMMDRPAINONVIOAR AHYRPNTRGRPGYHGRGIRYHRYHGRHNDGWNYPPLAFGAGAILGAIISQPR PVYRAPAGSHVQWCYSRYKSIRASNDTFQPYNGPKKCRSPYSR"

/complement(860..2029)

/gene="BMEI10553"

/complement(860..2029)

/gene="BMEI10553"

/EC_number="1.1.1.1"

/codon_start=1

/transl_table=1

/product="ALCOHOL DEHYDROGENASE"

/protein_id="AAL53795.1"

/db_xref="GI:17984726"

/translation="MTLITAKMNPFTVLFGCHIKELPAVLKAAGIEHPLFTDPGL AKLPYVASTLAVLDAAGVOYGVSDVKPNPVSNLHAGVKVFRGHGDIAPFGGSA LDGLKLIAPQAGQTRPVMDFEDIGDMTRADSNAPVIAVPTTAGTSEVGRAGVLT NESTHTKKVIPHMKLPVTVIADPELSTGMPPIITAGTMDALAHCLAYCAPGFHPM ADGIAVEGIRLVFENLPRAYANGDI EARNMMAAAGATAFOKGLGAIHSLSHPIG ALYDTHGNTNAPVFPVYVHNRAAIEKRIARLADYLGIEGDFGADAIKLRALIG VRTUPDLIKLQMONAKRELAEVADFTAGGNPVELTKDALKLNALEGR"

/complement(2130..3434)

/gene="BMEI10554"

/complement(2130..3434)

/gene="BMEI10554"

/EC_number="6.3.1.2"

/codon_start=1

/transl_table=1

/product="GLUTAMINE SYNTHETASE"

/protein_id="AAL53796.1"

/db_xref="GI:17984727"

/translation="MLACIVDMQGLIKGRFYGFVSGYDETHGCVLLADDIDME PVFGYAAAGWDKGYDGVNPKDLSTLRAPLWEKTAI VLCDVLDHHDHLSHSPRAVL KKQVORLHERGYRAYFASLEFYI FDETKSARAKRHEMETASPVQOQYVILHITRE EPLRAMRNLADAGIPVENSNGSGQQLNRYCKALENADRHVIMKNAMKEIAE AHGKCTIFMAKYDYARAGSSSHVHNSIWSADKEPLFDPKAPYTMTPLMRSWVAGQI KYATDYTVLAPYINSYKRFQAGTAPTKIMSDONRTAGFRLLCGEGTKGRIEGRIG KADINPYLAPALIAAGLKVDEKLEDEPFVFGDAYSAVRLKEIPYTLREAAALKGS AFLKPAFGEDVVNHYTHAHWEQIEYDRRVTDWELVGRFERY"

/complement(3548..4315)

/gene="BMEI10555"

gene

CDS


```

/codon_start=1
/transl_table=11
/product="AGR_C_3880p"
/protein_id="AAK87886.1"
/db_xref="GI:15157278"
/translation="MTVIONISPIDGSAAREAMSLEAARAAVSTARRAQKWARRP
LEDVQLVILGVARLENMAEAVPELAHMMGRPVRYGTFGRFNERNYVASIAAQT
APLVIESSGNFERRIEAGHGVFVIAPWPNYPYNTAVAPALMAGNTVYKHAQT
LLVGGERVAVGPEVDFVFNFLDGGTTSTLISEGLFNFNFTGSGVGGRAIERA
AAGFTGLGLELGGKDCPYVMEDADDAVDLMDGGTYNSGCCGCCGIERIYNNENLY
DEFVKSAYVAPQILVNVDHSMFEETFGPAIGIMVKNDDEALALANDSKYGLT
ASLTQDAARAARIGRIETGVFNNRADYLDPALCWTGKVTGKGTGSGSLVILGFQNL
RPSYHLKKVTK"
complement(3888..4334)
/gene="AGR_C_3881"
CDS
complement(3888..4334)
/gene="AGR_C_3881"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3881p"
/protein_id="AAK87887.1"
/db_xref="GI:15157279"
/translation="MILNGKSTMSDGLDTCVATWMSQGEVDSVSNLQSVR
AGDPVCVATVAMEGMLVAKRISETKSPQRWYDDFKREAIEVQPTADIETACF
LPQVHKDPIDRLITITAREHDLTIITRDRVILAYAGHVKTILAC"
complement(4297..4725)
/gene="AGR_C_3882"
CDS
complement(4297..4725)
/gene="AGR_C_3882"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3882p"
/protein_id="AAK87888.1"
/db_xref="GI:15157280"
/translation="MANTGQIDIKSWFSQOLKRGKGSQADLARFLNLDRAVTRMLNG
DRMSVEQDRIAEYIEIPVGDVALHRRGVAGVFSNNOTAYSEPAPSGRSGPEAGNV
STESRHPITGCKMGTITVMPDVLTKPVDLTPVDFEWGEKLYNE"
complement(4788..6152)
/gene="AGR_C_3883"
CDS
complement(4788..6152)
/gene="AGR_C_3883"
/notes="probable glnA4 protein - Mycobacterium tuberculosis
(strain H37Rv)"
/codon_start=1
/transl_table=11
/product="AGR_C_3883p"
/protein_id="AAK87889.1"
/db_xref="GI:15157281"
/translation="MTSYTFDALKMDVAEGRIDTVLACLVDMDQRLMGKRFQAEFFVE
SAPETHSCNYLATDEMETVPGYKSSWEKGYDYLKPDLSLTLKQVPLWGLGVALY
LCDVLDHHTHEEPHSPRALLKQVAKLEAMGLKAYNATELEFFLFDPQTPDAARASY
KDLNLASGYNEDHYFOTTKBEDVMRALRGLOAGIPVENSKEASPGQAEINRYA
EALTMADHRAIIKNATIEAMSKGKAVTFLAKWYNAAAGSSHHOSLWGLDGKPAFL
KDQGEKMSDMVHYAGLLAHASDITYFLAPYINSYKRPWAGTFPTKAIWSLDNPT
AGYRLCCAEKTRIERICRVGSDNPLVLAALLAGIDGLENKLELEPFPVGDYVGG
KDYREIPKTLRDATFADLGSKMLRQAFGDDVDVHYVHAARWEQSEYDRRVTDMEVARG
FERA"
complement(6218..7774)
/gene="AGR_C_3885"
CDS
complement(6218..7774)
/gene="AGR_C_3885"
/notes="probable transport protein PA4023 (imported) -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_C_3885p"
/protein_id="AAK87890.1"
/db_xref="GI:15157282"
/translation="MADWLVFLLSGEKGMDSSSGVSYYKKADASYFEKRLGSLYAGVW
SLWALGVAVISGHFSGWNFPGFTGGWGLVAGIIIAIMYLGITFSIAEMSPALPHT

```

```

GAAYSPARTAMGPGGFVTGLCENVEYVLPVAVVVTITAYVNSILGLDPAYSFPVMI
VFYAILANVFLGLEFSKVLITVITLISLAVVFFMISAIPIIDFSFALNIGVGPDG
KAVELPEGGSGFPFPGFVLATLPPAVMLFLAIQLPLAAESVDPKRMPKGIILG
MVTLMVSAFMVILLNPSLPGVGAFLHSSLSLEPLDGGFKAIYDGGVVLGLVALTGLI
ASPTLTYAQRGQIYSLSRAGYPTVLSITHSKYRTPYVANITGAIYGLAVMLVWFS
LGAEOGSIIGSVLLNNAVGAFFSYIMQAIISFILLKKNLPNIERPSPRSPGIPGALL
TIIIAIVTLIYQIDPNFTKGVLMVAVFAVAIYAFVAVGRHRLILSPSEBEFALEHQ
AAVAAAAKA"
complement(7956..8729)
/gene="AGR_C_3887"
CDS
complement(7956..8729)
/gene="AGR_C_3887"
/notes="similar to orf3 gene in Methylobacterium
extorquens"
/codon_start=1
/transl_table=11
/product="AGR_C_3887p"
/protein_id="AAK87891.1"
/db_xref="GI:15157283"
/translation="MTVRSRFFTEAEGKAVGVENAAKGDVLLVCEHASATIPQKYGT
LGLSADVLSHAAWDFCALAVARLLSEKPHATLVYQFSLVYDCNRPSPSPAMPVK
SEIYDIPGNFDLDEAREFARTSALYVPHDRVSEIIAEROAAGRKVVVVTIHSPTPVY
HGRPREVEIGILHDNDSRLADAMLAGAEGASLTVRNDPYCPEDGVTHTLRLHALPDG
LNMVIEIRNDLIANEGEQAIAGFLHELMKALSIEE"
8704..9780
/gene="AGR_C_3888"
CDS
8704..9780
/gene="AGR_C_3888"
/notes="hypothetical protein PA5506 (imported) -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_C_3888p"
/protein_id="AAK87892.1"
/db_xref="GI:15157284"
/translation="MKNRRTVMDPARKPAIPPMTKRILHIIASVKNLKFSLHFSI
DSAAGTDVYFSAPEDRKGRSLRTATVSDVIHAYDALTRESEKRLAESLLNGYV
SNAGSITTTAENAGVSTPTVARVQKLYGKVFQALHLOEALATISGPVAKHNRWA
TNAPGILHNRPADATIGNRLDGLDTPAFDFAARLLSDKRKSIYFVGRITGATA
EYFPTHMQVIRPKTTLMSNSSSAWPOYMLNMSAGDVLVIFDIRYHDMTTLAEVAKA
NQVQIILFTDQWTSVPVRAHALTFVRVRIEAPSADSSVTVFVVEALIEAVQNGTWDE
TKERNALGSLFEQTLRFRPERT"
complement(9799..10044)
/gene="AGR_C_3889"
CDS
complement(9799..10044)
/gene="AGR_C_3889"
/notes="hypothetical protein"

```

Alignment Scores:

Pred. No.:	2,01e-77	Length:	11253
Score:	942.50	Matches:	195
Percent Similarity:	63.11%	Conservative:	89
Best Local Similarity:	43.33%	Mismatches:	165
Query Match:	39.67%	Indels:	1
DB:	1	Gaps:	1

US-10-098-602a-2 (1-454) x AE008130 (1-11253)

Qy	4	ThrTyAspGluLeuAenAenLeuLeuAraGAsGlyLysIleAspThrValValLeuAala	23
Db	6140	ACATTCGACGACCTTAAATGGATGTCGCGAGGGGGCGATCGACATGCTTCTCGCGTCG	6081
Qy	24	CysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGly	43
Db	6080	CTTGTGCATCGAGGGGGCTCTGATGGGCAACGGTTCACGGCGGAATTCCTTCGTGAA	6021
Qy	44	LeuAspGlnLysIleSerIleSerThrPheValAlaValThrIleGluGlyIle	63
Db	6020	AGCGCCTTCAGGAAACCCATAGCTCAACTATGCTGCTCCACCCGACATGGAGATGAA	5961
Qy	64	AlaGlyGlyGlyTyrGluLeuSerValAspThrGlyTyrSerAspCysHisLeuCys	83
Db	5960	ACCGTTCCCGGTTTACAGTCTCGAGCTGGGAAAGGATATGCGCATATACGCTCAAG	5901

```
QY      84 AlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAlaIleSer 103
Db      5900 CCGGACCTTTCCGAGCTGCGGAAGTGGCGTGGCTGCAAGGTACGGCGCTTTTGTGTC 5841

QY      104 AsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeuMet 123
Db      5840 GAGTGCTGCACCATCACACCCATGAAGAAGTCCCGCATTCGCCGCGCGCTCTCTGAAA 5781

QY      124 GlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeuGlu 143
Db      5780 AAGCAGGTGGCGGCTGCGGAGCGATGCGGCTGAAGCGCTATATGCGCAGCACTAGAA 5721

QY      144 PheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisIleTrpLysAsnLeuLys 163
Db      5720 TTTCTTTCTGTTCACAGACCTTTTCGATGCGGCGCGCGCAGCGGTATATAAGACCTCAAT 5661

QY      164 ThrAlaGlnProHisHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThrPhe 183
Db      5660 CTCGCCAGCGGTATATGAGGATTACCATATCTTCCAGCAGCAAGGAGGAGGTGTC 5601

QY      184 MetArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHisPro 203
Db      5600 ATCGCGGCGTGGCAAGGCGCTGCAGGCTGCGGAATTCGCGTCGAGAATTCGAAGGCG 5541

QY      204 GluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThrMetAla 223
Db      5540 GAGGCGTCTCCCGGTGAGCGGGAATCAATGTGCGTTATGCGAGCGCTGACCATGGCT 5481

QY      224 AspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGlyMetVal 243
Db      5480 GACCGCAGCCCATCATCAGAACGCCACCAAGGAGATGCGCTGTCGAAGGCGCAAGCA 5421

QY      244 AlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIleHisMet 263
Db      5420 GTCACTGCTCTTCCCAAGTGAAGTCAATGCGCGCGGCGAGTTCATCGCATATCCACCA 5361

QY      264 SerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGluTyrGly 283
Db      5360 TCGCTCTGGAGCGCTGGAC---GGTAAGCGCGGCTCTCTCGACAGGATGCGCAACATGGC 5304

QY      284 MetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLysTyrValProGluAlaThr 303
Db      5303 ATGTCCGATGTGATGCGGCAATATGTGCGGGCTTCTGGCGCATCCGACGACATCACC 5244

QY      304 TyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPheAlaPro 323
Db      5243 TATTCTCGCGCATACATCAATCTCTACAGCGCTTCATGCTGCGCAGCTTTGGCGCG 5184

QY      324 ThrLysCysCysTrpAlaIleAspAsnArgThrSerAlaPheArgLeuCysAsnSerLys 343
Db      5183 ACAAGCGCATCTGGAGCTCGACACCGCACGCGCAGGATACCGCTTTGCGCGCGGAA 5124

QY      344 SerGluGlyIleAsnValGluLeuArgIleGlyAlaAspLeuAsnProTyrLeuAla 363
Db      5123 ACCAAGGCAATCGTATCGATGCTCGCGCGGCTCGGACCTCAATCCCTATCTCGCC 5064

QY      364 PheSerAlaIleIleAlaGlyIleSerGlyIleGluGlyLeuGluLeuProPro 383
Db      5063 ATGCGCGCTCTTCTGCGCGGCAATTCGCGGTATCGAAATAAATCTCGAATTCGAGCC 5004

QY      384 ProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSerLeuGln 403
Db      5003 GCTTTTGTGGTATGCTATGCGGTAAAGGAGCGGCGGAGATACCCCAAGCATTTACCG 4944

QY      404 AsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeu 423
Db      4943 GAGCGCAGCGCTTTCTGACCGGATCGAAATGTCTGCGCGCAGGCTTTGGCGACGAGTG 4884

QY      424 IleLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThr 443
Db      4883 GTCGATCACTACTGTCATGCGCAAGGTGGGAGCAGGAAGATACGACCGCGCGTGACC 4824
```

```
QY      444 AspTyrGluLeuAsnGlnGlyPheAsnArg 453
Db      4823 GACTGGAGTGGCACGCGGTTCGAGCGA 4794

RESULT 7
LOCUS   MTW003               13246 bp      DNA      linear      BCT 02-SBP-2002
DEFINITION   Mycobacterium tuberculosis H37Rv complete genome; segment 125/162.
ACCESSION   AL008883 AL123456
VERSION     AL008883.1 GI:3261490
KEYWORDS
SOURCE      Mycobacterium tuberculosis H37Rv
ORGANISM    Mycobacterium tuberculosis H37Rv
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE   1
AUTHORS     Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
            Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
            Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
            Connor, R., Davies, R., Devlin, K., Feltwell, T., Genes, S.,
            Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
            Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
            Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
            Squares, S., Squares, R., Sultston, J.E., Taylor, K., Whitehead, S. and
            Barrell, B.G.
TITLE       Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
JOURNAL     Nature 393 (6685), 537-544 (1998)
MEDLINE    98295987
PUBMED     9634230
REFERENCE   2 (bases 1 to 13246)
AUTHORS     Parkhill, J.
TITLE       Direct Submission
JOURNAL     Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
            On Jun 27, 1998 this sequence version replaced gi:2612801.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes.
            Implemented in TPAse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION: In some cases we may not have predicted the correct
            initiation codon. Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon) if this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
                     1..13246
                        /organism="Mycobacterium tuberculosis H37Rv"
                        /mol_type="genomic DNA"
                        /strain="H37Rv"
                        /db_xref="taxon:83332"
                        <1..>13246
                        /note="fragment designated v003. Does not represent a
                        physical clone"
misc_feature         /gene="Rv2857c"
                     /complement(60..836)
gene                /gene="Rv2857c"
                     /complement(60..836)
CDS                 /gene="Rv2857c"
                     /note="Rv2857c, (MTW003.03c), len: 258 aa. Probable
                     dehydrogenase, similar to many eg. P94681 toluenesulfonate
                     zinc-independent alcohol dehydrogenase (tsac) from
                     Comamonas testosteroni (252 aa), fasta scores; opt: 515
```

z-score: 632.1 E(): 4.7e-28, 38.0% identity in 250 aa overlap. Also similar to many M. tuberculosis dehydrogenases eg. MTCY39.16c (38.3%identity in 248 aa overlap)"

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2857c"
/protein_id="CAA15519.1"
/db_xref="GI:2612802"
/gb_xref="SPTREMBL:O33339"
/translation="MMDLSORLAGRVAIVTGGSGIGLAAGRRMRAGGATVVGVDVDF
EAGGNADELGLFVFDTDCEDAVNGLVFGAASYGRIDIAFNAGISPEENLIEN
TELAQRQVDNLKSVTLQRCRAALRHMLAKGSIVNTASVFAVMGSAFSQSITYTAS
KGVLAEIRGEVGFQAFQIRVNALPCPVNTPLLQLFAPKNPERAARMVMHVPLQRF
APEPILAAAFVFLASDDASTFTSLFDVGDISAYVTPL"
complement(833..2200)
/gene="aldC"
complement(833..2200)
/gene="aldC"
/note="RV2858c, (MTW003.04c), len: 455 aa. aldC,
Probable aldehyde dehydrogenase, aldC; similar to many eg.
FEAB ECOLI P80668 phenylacetaldehyde dehydrogenase (499
aa) _fasta scores; opt: 1074 z-score: 1239.2 E(): 0, 42.2%
identity in462 aa overlap. Contains PS00687 Aldehyde
dehydrogenases glutamic acid active site and PS00070
Aldehyde dehydrogenases cysteine active site. Also similar
to many M. tuberculosis dehydrogenases eg. MTCY369.13
(38.1% identity in 467 aa overlap)"
/codon_start=1
/transl_table=11
/product="aldC"
/protein_id="CAA15520.1"
/db_xref="GI:2612803"
/gb_xref="SPTREMBL:O33340"
/translation="MSUTOLINPAATESVLASVDHTDANAVDDAVQRAARAORPWRLA
PAQRAUGRAFAAAVQHLEDALEAVEANGHPIVSAEWAGHVRDVLFATYASPRLT
SGROI PVAGGVDTVFNEPMGVQVITPNFPMWIASWAIPAALGAANVLAVPKAELTP
LTWRKLGLAEVAGDELDLVLPGKGTVGGERFVTHPDIKIVFTGSTEVGRKRWG
AAAOVRLTELGGKSANI VPHDCDLERAATTAPAGVFDNAGODCCARSRLIQVSRY
DRFWELLEPAVHSIIVGDGPGRATTEMGPLVSRAHRDKVAGVVPDPADPAVAFRGTA
GFVEPPTVLTFRGDRGRTVTDIFGPFVVVLTLPDDEADAISLANDTAYGLSGSIWTDDL
SRALKAVAKAVESGNLSVNSVSNFRNTPFFGGFKOSGVRELGPDLQFTETKNNVILA
VGEEM"

RBS
complement(839..843)
/gene="aldC"
/note="possible RBS upstream of Rv2857c"
complement(1400..1435)
/gene="aldC"
/note="PS00070 Aldehyde dehydrogenases cysteine active
site"

misc_feature
complement(1496..1519)
/gene="aldC"
/note="PS00687 Aldehyde dehydrogenases glutamic acid
active site"
complement(1297..3123)
/gene="Rv2859c"
complement(2197..3123)
/gene="Rv2859c"
/note="RV2859c, (MTW003.05c), len: 308 aa. Unknown;
similar to E. coli hypothetical gene TR:P76038
(EMBL:AB000228) (258 aa), fasta scores; opt: 326 z-score:
354.5 E(): 1.4e-12,30.7% identity in 238 aa overlap.
Contains three 17 aa repeats at the N-terminus very
similar to those in other M. tuberculosis proteins e.g.
Y130 MYCU Q10699 putative DNA polymerase c149.30 (393:
aa)"

misc_feature
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2859c"
/protein_id="CAA15521.1"
/db_xref="GI:2612804"
/gb_xref="SPTREMBL:O33341"
/translation="MDLSASRSKSDGGDPLRPASPRLRSPVSDGGDPLRPASPRLRSPVSD

Db 4453 TTGGCGTGGACCGAGTTGGACGACTGGTCCGCGCGGTGACGTCGACACCGCTCATGTC 4394
 QY 23 AlaCysValAspMetGlnGlyArgLeuMetGlyLeuArgLeuThrGlyArgHisPheLeu 42
 Db 4393 GCCTTACCACATGACGGCCGGCTGGCCGCAACCGATATCGGCGCGCATTTTCGTC 4334
 QY 43 ---GlyLeuAspGlnIleValSerHisPheValThrPheValThrLeuGlu 61
 Db 4333 GACGACATACCCCGCGCGTGGAGTCTGTCAGTTATCTGTCGCGCGTGGACGTCGAC 4274
 QY 62 GlyIleAlaGlyGlyThrGluIleSerValAspThrGlyThrSerAspCysHis 81
 Db 4273 CTGAACACGCTGCCGCTATGCGATGCGCAGTTGGACACCGGCTACGCGCATGGTG 4214
 QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTyrSerGluGlyAlaValLeuAla 101
 Db 4213 ATACGCGCGACTGTCACCTCTGCGGCTGATTCCTTGGTACCGGAACGCGCGTGGTG 4154
 QY 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIle 121
 Db 4153 ATCGCCGAC---CTGGTCTGGCGCGGACGAGTCCGCTCTCGCGCGCAGCAT 4097
 QY 122 LeuMetGlnGlnIleGluArgLeuAlaAsnLeuLeuLeuGlyLeu-----Phe 138
 Db 4096 CTGCGCGCTCAGCTCGATCGG-----CTCAAGCGCGCGGACTGGTTCGCGCATGTG 4046
 QY 139 AlaSerGluLeuGluPheAsnLeuPheAsnGluThrTyrLeuSerAlaSerGlnIleHis 158
 Db 4045 GCACCGAGTGGAGTTTCATGTTTCGACCGCTATCGCCAGGATCGGCGACGCGGG 3986
 QY 159 TrpLeuAsnLeuLeuThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSer 178
 Db 3985 TATCGCGGCTGACCGCGCGGACGACTACACATCGACTACCGATATGGCATCTCTCG 3926
 QY 179 GlyIleGluThrPheMetArgSerValArgAsnLeuLeuGluAlaGlyIleLeuMet 198
 Db 3925 CGGATGAGCGGTTGTCGCGCATCGGTTGGTATGGCGGTGGTCTCGCATTC 3866
 QY 199 GluAlaThrHisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp 218
 Db 3865 GAGCGGTCAAGCGGATCAACATGCGCGGACGAGATCGGGTTTCGTTACGACGAG 3806
 QY 219 ProLeuThrMetAlaAspArgHisIleAlaLeuHisGlyValArgGluMetAlaGlu 238
 Db 3805 CGCGTGTGCTACCTCGACACCATCGCATCTACAGAACGCGCCAGGAATCGCGAC 3746
 QY 239 GlnSerGlyMetValAlaThrPheMetAlaLeuSerSerThrAlaLeuGlyLeuAsnAla 258
 Db 3745 CAGCAGCGCAAGAGCCTTACGTTTCATGCGCGAATACGATGAACGCGAA---GGTAATAGC 3689
 QY 259 CysHisIleHisMetSerLeuGlnAspAlaGluThrGluLeuAsnAlaPheTyrAspGln 278
 Db 3688 TGTCAATCATGCTGCTGCTGCGTGGCGACGATGGTTCGCG---GTGTTTGGCGACAGT 3632
 QY 279 AsnAspGluThrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLeuTyr 298
 Db 3631 AACGGCGCGCACGCGCATGCTCGATGTTCCGCGAGTTCGTCGCGCGCAGTTGGCCACG 3572
 QY 299 ValProGluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrIleArgLeuGlnPro 318
 Db 3571 TTGCGCGGAATTCACGCTGTGCTATGCGCGCACCACTTAACCTTACAGCGATTTGCCGAT 3512
 QY 319 LeuThrPheAlaProThrLysCysCysTyrTrpAlaIleAspAsnArgThrSerAlaPheArg 338
 Db 3511 AGCAGTTTCGCGCGGACGCGGCTGGTGGGCTGGACAAATCGACCTCGCCCTCGCGG 3452
 QY 339 LeuCysAsnSerLysSerGluGlyIleAsnValIleLeuArgIleGlyGlyAlaAspLeu 358
 Db 3451 GTGGTT---GGCCACGGCAAAACATCCGGTTCGAATCGCGGTTCGCCGCGGTGATGC 3395
 QY 359 AsnProTyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLys 378
 Db 3394 AACCACTACCTGGCGGTGGGCTCTCATTTGCTGGAGGGTTGTACGTATCGAGTACGAGGGG 3335

QY 379 LeuGluLeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPhe 398
 Db 3334 CTTACGTCGCCGAGCCCTGTGTCGCAACGCGCTACCAAGCGCGCATGTCGAACGCGTG 3275
 QY 399 ProAsnSerLeuGlnAsnAlaThrHisLeuLeuLeuGlySerLysMetLeuAsnLysThr 418
 Db 3274 CCGGTTACGTCGCCGACGCCGCGGTCTGTTGTCGAGATTCTGCGCTGGTGGCGAGCG 3215
 QY 419 PheGlyGluLeuLeuIleLeuHisTyrValAsnAlaAlaAsnValGluLeuAsnGluPhe 438
 Db 3214 TTGCGCGAGGATGTTGTCGCGCACTACCTGAACACGCGCTGTGAGCTGGCGGCGTTC 3155
 QY 439 SerLysGlnValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
 Db 3154 AACCGCGCGCTACCGATTGGAGAGGATACGTGGATTTCGAGCGC 3110
 RESULT 8
 AE007117/c 14715 bp DNA linear BCT 27-APR-2001
 LOCUS Mycobacterium tuberculosis CDC1551, section 203 of 280 of the
 DEFINITION complete genome.
 ACCESSION AE007117 AE000516
 VERSION AE007117.1 GI:13882699
 KEYWORDS
 SOURCE Mycobacterium tuberculosis CDC1551
 ORGANISM Mycobacterium tuberculosis CDC1551
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 REFERENCE 1 (bases 1 to 14715)
 AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
 Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
 Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
 Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
 Gill, J., Mikula, A. and Bishai, W.
 TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 14715)
 AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
 Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
 Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
 Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
 Gill, J., Mikula, A. and Bishai, W.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1. .14715
 /organism="Mycobacterium tuberculosis CDC1551"
 /mol_type="genomic DNA"
 /strain="CDC1551"
 /db_xref="taxon:83331"
 /note="clinical strain"
 complement (148..924)
 /gene="WT2925"
 complement (148..924)
 /gene="WT2925"
 /note="similar to GB:AE000512; identified by sequence
 similarity; putative"
 /protein_id="AAK47250.1"
 /codon_start=1
 /transl_table=11
 /product="oxidoreductase, short-chain
 dehydrogenase/reductase family"
 gene
 /protein_id="AAK47250.1"
 /db_xref="GI:13882700"
 /translation="NMDSQRLAGRVAVITGGSGIGLAAGRMRAEGATIVVGDV
 EAGAADELGLFVPTDVCDEDAVNLGFDGAAETVGRIDIAFNAGISPPENLLEN
 TELAAQRQVDVNLKSVLCCRAALRHVYACKSVNTASVAVMGSAISQISYAS
 KGVLAWSRELGVQARQIRVNLCPGVNTPLLQSLFAKPERAARMVHVLGRF
 ASPDEIAAFAVLASDASFITATFLVDGGISSAYVTPL"
 complement (921..2288)
 CDS
 /gene="WT2925"
 complement (148..924)
 gene

CDS	/gene="MT2926" complement (921. .2288) /gene="MT2926" /note="similar to PID:2228233; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="aldehyde dehydrogenase family protein" /protein_id="AAK47251.1" /db_xref="GI:13882701" /translation="MSTQLINPATEEVLASVDHADANAVDQVRAAQRWARLA PAQRAGURAPAAVAHLELALEAVNSGHPIVSAWEAGHAGVNDVLAFAASPERL SGRIQVAGGVDDVTFNPMGVGVITPMNFMVIAWSAIAAPALAGNAVLVKPAELTP LTMRLGLAEAGLEDDLLQPKGVVGERFVTHPDIRKI VFTGSTEVGKVMAG AAQVKRVTLELGSANI VPHDCDLERAAATPAGVFDNAGQCCARSRLVQKRSVY DRFVELLEPVAHSIVGDPGSRATEMGLVSRHRDKVAGVVPDPAVAPRGTPAGR GFWPPPTVLTFRGDRVTDSIFGPVVVVLTFDDDEADAI SLDATAYGLSGSIWTDL SRARVAVESGNLSVNSHGRFNTFPFGFKOSGVGRELGPAPLQFTETKNVFA VGEEN" complement (2285. .3085) /gene="MT2927" complement (2285. .3085) /gene="MT2927" /note="similar to GP:3581870; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glutamine amidotransferase, putative" /protein_id="AAK47252.1" /db_xref="GI:13882702" /translation="MSGDGLPLRPASPLRSLPGASRPVVGLTAYLEQVTRGVNDIPA GYLPADYPEGITMAGGVALLPPQVPESVGCVLDSLHALVITGGVDLPAAVGOEP HPATDHPRGDAWEFALLRGALORGMPVLGICRGTOVNLVAGTGLHOHLPLDILGHS GHRAGNVFTRLPVHTASGTRLAELIGESADVPCYHQAIIDVQGLVSVAVDVGVI EALPLDGTFLVAVQWHPKSLDLRLFKALVDAAAGYAGRQSAEPR" complement (3138. .4511) /gene="MT2928" complement (3138. .4511) /gene="MT2928" /note="similar to SP:P36205 GB:X60160 PID:48181. GB:AE000512; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glutamine synthetase, putative" /protein_id="AAK47253.1" /db_xref="GI:13882703" /translation="WTGPGSPPLAWTELERLVAAAGVDVTIVATDMQGLAGKRISG RHFVDDIATRGVECCSYLLAVDVLNTPVGYAMASWDTGYGDMWMTDLSLRLIPWL PGTALVADLVNADGSEVAVSPSILRRQLDLKARGLADVATELEFIVFDQPYRQA WASGYRGLTPASDNIYDAILASSMEPLLRDIRLGWAGAGLRFAVKGEKCMQOQBI GFYDEALVTCNDHAIYKNGAKETADQKSLTFMAKYDEREGNSCHIHVSLRGTDGS AVEADNSPGHMSMFRSFVAGOLATLREPTLCYAPTINSYKRPADSFPAPTALANGL DNRTCALRVGHNQIRVECRPGVDNQYLAVALTAGGLYIERGLQPEPCVGNV YQGADVRLPVTLADAAVLFDLSALVREAFGEDVVAHYLNNARVELAAFNAAVTDWR IRGPERL" complement (4671. .5528) /gene="MT2929" complement (4671. .5528) /gene="MT2929" /note="similar to GB:D26562 SP:P07906 GB:M15106 PID:146727 PID:473823; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="methionine aminopeptidase" /protein_id="AAK47254.1" /db_xref="GI:13882704" /translation="MPSRTALSPGLSPTREVPNWIARPEYVVGKPAAQEGSEPWQTP EVIEMRVAGRIAGALAEAGKAVPGVTTDELDRIAHEYLVDNGAVPSTLYKGFPK SCCTSLENI CHIGIPDSTVITDGI VINDIVTAYIGVGHGDTNATFPAGDVADHERLLV DRTEATMRAINTVQGRALSIVGRVIESYANRGYNVRDFTGCHGTTFHNGLVVL HYQPAVETINQKPGMTFTIEPMINILGALDYIEIWDGWTVTVTOKRKKTAQFHTLLVTD TGVILITCL" complement (5570. .6154)	CDS	/gene="MT2930" complement (5570. .6154) /gene="MT2930" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK47255.1" /db_xref="GI:13882705" /translation="MTETGDMVALRVSDADRNGTMRLLHNAVALGILINIDPEQORS RVSPACTRSLDGLVGLDLPREGALVTSAADRVELRWAGSLKRGHEWIVTRLLAVRR LGSLELDLVKARFAPVVVIELDMFMFGLSLEVRLPNGASATDDVEVYVGSASDRRKKDA PAGCTPHVITLGRMVCVGVVVKGPRRALLRRHG" 6312. .6506 /gene="MT2931" 6312. .6506 /gene="MT2931" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47256.1" /db_xref="GI:13882706" /translation="MLSNLWLRQAGLRQLEAQRQRLTAQELREFFPASRPDETGAEP DWAHLQVMAESRRRGLPAP" 6503. .6883 /gene="MT2932" 6503. .6883 /gene="MT2932" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47257.1" /db_xref="GI:13882707" /translation="MIFVDVTNFMVYAVGRDPLRMPAREFLEHSLHQDLRLVTSAEAM QELNAYVPVGRNSTLSDALTALREALTEIWPVEAADVAHARTLHRRHPLGLGARDLLHL ACQCRGVTRIKTFDHTLASAFRS" complement (6965. .8821) /gene="MT2933" complement (6965. .8821) /gene="MT2933" /note="similar to GP:1666696; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="penicillin-binding protein, putative" /protein_id="AAK47258.1" /db_xref="GI:13882708" /translation="MSRPVSPQFLTSATWTKTTLASATSGLLLLAVVAMSGCTPRP QGPQAEKFFAALAGDTPASAAQLSDNPNEAREALNAWAGLQAHLDAQVLSAKYA EDTGTVAFRSWHLPKDRITWYDQGLKMARDEGRWVRWTTSGLHPLKLGHTQTFALRA DPRRASVNVGGTDVLPGLYLHYSLDAGQAGRELFGTAHVAVGALHPDDTLDNDPO LLAEQASSSTQPLDLVTLHADDSNRVAAAIQOLPGVITTPQALLPTDKRHPAVALND VKCAVDELDKAGRWVSVYQNGVDVSVLHEVAPSPASSVITLDRVQVNAQAHAVN TRGKAMIVVVKPSTGEILAIQONAGADADGPVATGLYPGSGTFKMITAGAAVERDL ATPETLLGCGEIDIGHRTIPNYGGDLGVVNSRAPASSCNTTFAELSRLLPRLGLT QAARTVPGILDYQDGTITVTGVSPTVDLAERTBEDGFGQGVKVASFPFGNALVATVA AAGKTPVQLIAGRPATVEGDATPI SQMIDALRPMRLVVTNGTAKETACGCEVFGKT GEAEFFGGSHWFPAGYRGDLAFALISLVGGSSSEYAVRMTKVMFESLPPGYLA" complement (9006. .9716) /gene="MT2934" complement (9006. .9716) /gene="MT2934" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK47259.1" /db_xref="GI:13882709" /translation="MPPHEHRNPEVLFPSFRSRPADRSQSWLPDSGRAPVGBLLSAV DVRADVDAQDHYRCVLVNPVQOAVRSATRAERAGLAPKGLAHPQGLARQIAEREPD HCRDSRWQLVEVSTRGCGPHGRHRSVGSASRDABFGADLVFAVGAAGNVGVGFS	gene
-----	---	-----	---	------

```
/db_xref=taxon:233413"
complement(99.875)
/locus_tag="Mb2882c"
```

```

CDS
complement(99..875)
/locus_tag="Mb2882c"
/EC_number="1.-.-.-"
/notes="Mb2882c, -, len: 258 aa. Equivalent to Rv2857c.
len: 258 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 258 aa overlap). Probable short-chain
dehydrogenase/reductase (EC 1.-.-.-), highly similar to
various dehydrogenases e.g. O88068|SCI35.33c PROBABLE
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
identity in 253 aa overlap); O91376|PA1649 from
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
(253 aa), FASTA scores: opt: 569, E(): 2.1e-28, (39.2%
identity in 255 aa overlap); O9EX74|MLHA SDR-LIKE ENZYME
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);
etc. Also similar to many Mycobacterium tuberculosis
dehydrogenases e.g. FAE3|RV2002|MT2058|MTCY39.16c
PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
DEHYDROGENASES/REDUCTASES (SDR) FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE SHORT-CHAIN TYPE
DEHYDROGENASE/REDUCTASE"
/protein_id="CAD96569.1"
/db_xref="GI:31619629"
/translation="MMDLSQRLAGRAVITGGSGIGLAAGRRMARGATIVGSDVDV
EAGGAADSLGLFPVTDVCDDEAVNGFLDGAAGTYGRIIDIAFNAGISPEDNLIN
TELAWQVDVNLKSYLCRCALRHMLVAGSIVNTASFAVMGATSQISYLTAS
KGGVLMSRELGVQFAGIRVIRNCPGVNTPLQLFAPKNPARRMVRHVPVLRGF
AGPEIDAAVAFASDSDASFTTASTFLVDGIGISSAYVTPL"
complement(872..2239)
/genes="aldC"
/locus_tag="Mb2883c"
complement(872..2239)
/genes="aldC"
/locus_tag="Mb2883c"
/EC_number="1.2.1.3"
/notes="Mb2883c, aldC, len: 455 aa. Equivalent to Rv2858c,
len: 455 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 455 aa overlap). Probable aldC.
aldehyde dehydrogenase (EC 1.2.1.3), similar to many e.g.
O88069|SCI35.34c PUTATIVE ALDEHYDE DEHYDROGENASE from
Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,
E(): 6.4e-109, (64.5% identity in 448 aa overlap);
O9FAB1|ALDH|BT-ALDH ALDEHYDE DEHYDROGENASE from Bacillus
thermoovorans (497 aa), FASTA scores: opt: 1157, E():
2.1e-64, (44.3% identity in 458 aa overlap); O33455|CYMC
P-CUMIC ALDEHYDE DEHYDROGENASE from Pseudomonas putida
(494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15%
identity in 452 aa overlap);
P40047|DHAS YEAST|ALD5|ALDH5|ALD3|YER073W ALDEHYDE
DEHYDROGENASE from Saccharomyces cerevisiae (Baker's
yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60,
(38.55% identity in 459 aa overlap);
P80668|FEAB ECOLI|PADA|MAOB|B1385 PHENYLACETALDEHYDE
DEHYDROGENASE (EC 1.2.1.39) from Escherichia coli strain
K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2%
identity in 462 aa overlap); etc. Also similar to many M.
tuberculosis dehydrogenases e.g. P71823|Rv0768|MTCY369.13
(489 aa), FASTA score: (38.1% identity in 467 aa overlap).
Contains P500687 Aldehyde dehydrogenases glutamic acid
active site and P500070 Aldehyde dehydrogenases cysteine
active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES
FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"
/protein_id="CAD96570.1"
/db_xref="GI:31619630"
/translation="MSTQLINPATEVELASVDHTDANAVDDAVQRAARAORWARLA
PAQRAAGLRAFAAAVQAHDLAALAEVANSCHPTVSAEWAEGHVRDVLAFYAASPERL
SGRIIPVAGGVDVTFNEPMGVGVITPMFPMVIASWAIAPALAAAGNAVILVKPAELTP
LTTMRLGLAVEAGLQDLQVLPGKTVVGERFVTHPIRKIVITGSTEVGKRVAMAG
AAQVRKVTLELGGKSNKIVFHDCLDLEAATTPAGVDFNAGODCCARSTLQVRSVY
DRMELLEPAVHSIVVCGDPCGRATEMGLVSRHRDKAVGVDPDPAVARGTAPAGR
GFWPPFVLTPKGRDVTNDEIFGPPVVVNLTFDDEADAISLANDATAYGLSGSIWTDLL
SRALRVARAVESGNLSVSSHSSVFNFPPFGFKQSGVGRGLGDPAPLOFTETKNVFIA
VGSEM"
complement(2236..3162)
/locus_tag="Mb2884c"
complement(2236..3162)
/locus_tag="Mb2884c"
/EC_number="2.-.-.-"
/EC_number="6.3.5.-"
/notes="Mb2884c, -, len: 308 aa. Equivalent to Rv2859c,
len: 308 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 308 aa overlap). Possible
amidotransferase (EC 6.3.5.- or 2.-.-.-), equivalent (but
longer 58 aa) to Q9CBU9|MLI573 POSSIBLE AMIDOTRANSFERASE
from Mycobacterium leprae (249 aa), FASTA scores: opt:
1226, E(): 3e-64, (71.55% identity in 239 aa overlap).
Also similar to other amidotransferases and hypothetical
proteins, but shorter in N-terminus e.g. O88072|SCI35.37
HYPOTHETICAL 25.3 KDA PROTEIN from Streptomyces coelicolor
(242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65%
identity in 235 aa overlap); AAK79730|Q97188|CAC1764
PREDICTED GLUTAMINE AMIDOTRANSFERASE from Clostridium
acetobutylicum (241 aa), FASTA scores: opt: 458, E():
1.6e-19, (32.95% identity in 246 aa overlap);
AAK75091|Q97QV9|SPI089 GLUTAMINE AMIDOTRANSFERASE CLASS I
from Streptococcus pneumoniae (229 aa), FASTA scores: opt:
431, E(): 5.6e-18, (34.75% identity in 236 aa overlap);
etc. Contains three 17 aa repeats at the N-terminus very
similar to those in other Mycobacterium tuberculosis
proteins e.g. Q10699|YY30 MYCTU|RV2090|MT2151|MTCY49.30
PUTATIVE 5'-3' EXONUCLEASE RV2090 (EC 3.1.11.-).
/codon_start=1
/transl_table=11
/product="POSSIBLE AMIDOTRANSFERASE"
/protein_id="CAD96571.1"
/db_xref="GI:31619631"
/translation="MDLSASRSGGDPRLRASPRLSPVSDGGDPLRASPAPRSPVPS
DGGDPLRASPRLSPVGLTAYLEQVTVGMDIAGYLPADYFEGITMAGS
VALLPPQVPDPESVGVCLDSHALVITGGYDLDPAAYGQEPHPADHPGRDAPWF
ALLRQALQGMGPVLGICRGTVNLVALGTLHQLPDLILGSHGRAGNVFTRLPVHT
ASCTRLAEIGESADVPVCHQALDQVGEGLVSAVDVGVIEALEPLPGDTFFLAVQM
HPEKSLDLRLFKALVDASGYAGRSQAEPR"
complement(3143..4516)
/genes="glnA4"
/locus_tag="Mb2885c"
complement(3143..4516)
/genes="glnA4"
/locus_tag="Mb2885c"
/EC_number="6.3.1.2"
/notes="Mb2885c, glnA4, len: 457 aa. Equivalent to Rv2860c,
len: 457 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 457 aa overlap). Probable glnA4,
glutamine synthetase class II (EC 6.3.1.2), similar to
many glutamine synthetases e.g. O88070|SCI35.35c from
Streptomyces coelicolor (462 aa), FASTA scores: opt: 1947,
E(): 8.2e-120, (64.15% identity in 452 aa overlap);
Q98H15|MLL3074 from Rhizobium loti (Mesorhizobium loti)
(465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7%
identity in 452 aa overlap); O98EM0|MLL4187 from Rhizobium
loti (Mesorhizobium loti) (456 aa), FASTA scores: opt:
699, E(): 4.6e-38, (33.5% identity in 454 aa overlap);
Q9CDL9|GLNA from Lactococcus lactis (subsp. lactis)
(Streptococcus lactis) (446 aa), FASTA scores: opt: 633,
E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.
Also similar to three other potential glutamine synthetases
in Mycobacterium tuberculosis:
Q10378|GLN2 MYCTU|GLNA2|RV2222c|MT2280|MTCY190.33c|MTCY427
Q10378|GLN2 MYCTU|GLNA2|RV2222c|MT2280|MTCY190.33c|MTCY427
. 03c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score:
(31.1% identity in 453 aa overlap); Rv1878|glnA3 and
```


JOURNAL

Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1BA E-mail: sdb@anger.ac.uk
On or before Oct 29, 2002 this sequence version replaced
gi:20520886, gi:20520900, gi:20520927, gi:20520938, gi:20520939,
gi:20520834, gi:20520914, gi:20520916.

FEATURES

source

Location/Qualifiers

1..277000
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"

/strain="A3(2)"

/db_xref="taxon:100226"

226..2993

/gene="SC01445"

/note="synonym: SCL6.02"

226..2993

/gene="SC01445"

/note="SCL6.02, hypothetical protein, len: aa; similar to various hypothetical proteins, e.g. TR:CBAS2958 (EMBL:AL109950) Streptomyces coelicolor hypothetical 25.2 kD protein, SCJ4.24c, 242 aa; fasta scores: opt: 424 z-score: 482.7 E(): 1.8e-19; 39.9% identity in 218 aa overlap and to TR:O9BWF2 (EMBL:AL451182) Streptomyces coelicolor hypothetical protein SCK13.32c, 263 aa; fasta scores: opt: 711 Z-score: 779.3 E(): 8.9e-36; 49.407% identity in 253 aa overlap"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein SCL6.02"

/protein_id="CAB76867.1"

/db_xref="GI:7209205"

/translation="MTTILVTGGTGLRLVTERLTCGHEVRVLSRHSFPYAVDLRE GSGDLAALGADVTVHCATTQCGDERSAANLAAARACAGLHVTISIVGVDRVPL GYRSKVAEKLVAESGIGTWLRTQDFDLVQMLRALAKLPVWLPAHASDPQVPL AEVADRLAEALQAQACAGCPVDMGCPVETASLARAYLGATGRRRAVNVVPLAGRTYR AFRAGHLAEHAGEGTTFPEYLAGRGHGGRRRG"

1064..1432

/gene="SC01446"

/note="synonym: SCL6.03"

1064..1432

/gene="SC01446"

/note="SCL6.03, possible integral membrane protein, len: 122 aa. Contains possible hydrophobic membrane spanning regions"

/codon_start=1

/transl_table=11

/product="possible integral membrane protein"

/protein_id="CAB76868.1"

/db_xref="GI:7209206"

/translation="MDPAAARAVLEVALCPAALHEHLRLDFGGAGLALAVMLCAAV ATWERLVLTPVACPDSSVARLLFHARHLEPLSAGGAAGFMPLGTAVLLPTVLGWP AADGTAEPGSACTGRRPG"

complement(1454..2653)

/gene="SC01447"

/note="synonym: SCL6.04c"

complement(1454..2653)

/gene="SC01447"

/note="SCL6.04c, possible ROK-family transcriptional regulatory protein, len: 399 aa; similar to TR:CBAS71832 (EMBL:AL13862) Streptomyces coelicolor putative ROK-family regulatory protein SCB4A.27, 409 aa; fasta scores: opt: 938 z-score: 965.1 E(): 0; 42.5% identity in 398 aa overlap and to SW:XYLR ANATH (EMBL:269782) Anaerocellum thermophilum xylase repressor XylR, 399 aa; fasta scores: opt: 327 z-score: 341.8 E(): 1.3e-1; 22.2% identity in 397 aa overlap. Contains Pfam match to entry PF00480 ROK, ROK family and possible helix-turn-helix motif at residues 26..47 (+2.84 SD)"

/codon_start=1

/transl_table=11

/product="putative ROK-family transcriptional regulatory protein"

/protein_id="CAB76869.1"

/db_xref="GI:7209207"

/translation="MPASPSTARAINDRLALHLIQREGPLTAGOLKLTGLSRPTVAD LVERLTASGLIEVAGSEQRGNKLYGVAERAYLAALDVTEGVAVLVSLGLGR VLAESVPIIDHSGSTGPAVQAVSLVERAKEAGADRLHTVGIGAPGLIDPAGGELRD STGLPEWHRLMALQLEKFEARVGVENETNLAAEQRDGAARDTFTVLLWLGIGI GAAVLDGTLRRGVSGGAGIEPVPGTAGLSDTDCDGGFSVAGAAVYVLDPGC GVTAPAGPEPFAAALVRRSASGSDPAARFLDALADRVVLGAAVYVLDPGC LVLAGETAGAGADALADRVQHLRTRMSPLATEVRPSTLGGGAVLRGALLTARDRAQDD LFAPPER"

complement(1934..2104)

/gene="SC01447"

/note="Pfam match to entry PF00480 ROK, ROK family, score 45.80, E-value 1.5e-12"

complement(2662..2667)

2753..2757

2766..3977

/gene="SC01448"

/note="synonym: SCL6.05"

2766..3977

/gene="SC01448"

/note="SCL6.05, probable transport protein, len: 403 aa; similar to TR:AAF12043 (EMBL:AE002079) Deinococcus radiodurans transport protein, putative DR3502, 471 aa; fasta scores: opt: 687 z-score: 761.8 E(): 0; 39.0% identity in 359 aa overlap and to SW:MOSC RHIME (EMBL:U17071) Rhizobium meliloti membrane protein MosC, 407 aa; fasta scores: opt: 496 z-score: 552.3 E(): 2.4e-23; 31.0% identity in 378 aa overlap. Contains match to Prosite entry PS00217 Sugar transport proteins signature 2. Contains also possible hydrophobic membrane spanning regions"

/codon_start=1

/transl_table=11

/product="putative transport protein"

/protein_id="CAB76870.1"

/db_xref="GI:7209208"

/translation="MNTGTGVREVRHARYAAVFAVHGSVTGSFATRPVWIOQHAG VSAGQLGALAFPALGASDVAANLGVETRLDKSINSLGLHGMSTGLIGSAAGTLAA TLCLALFVYGATAGMSDVAANLGVETRLDKSINSLGLHGMSTGLIGSAAGTLAA HLCTDARTHLLAAAVLTGVLVACTVMDLOPAEDEDPPRFPALPKSALLIGAVGF CAVFAGASLDMSAVYLRDELETSSAGLAACTTGFTLTMARLAGDKVVDRFQAVRT VRVSGVLATIGLLVVVGGHPAVAMTGFALMGIGIAVVPCLCAAAGAGSNPSLAIA GVATITTSGLVAPSAIGGLAQATSLVWSFGLVTLACCLVFAVRLRAGDRNRPKIS PPVAEVRGQS"

3093..3170

/gene="SC01448"

/note="PS00217 Sugar transport proteins signature 2"

complement(3937..4395)

/gene="SC01449"

/note="synonym: SCL6.06c"

complement(3937..4395)

/gene="SC01449"

/note="SCL6.06c, hypothetical protein, len: 152 aa;

similar to SW:YBGC ECOLI (EMBL:M16489) Escherichia coli hypothetical 15.6 kD protein in CydB-TolQ intergenic region ybgC, 134 aa; fasta scores: opt: 140 z-score: 193.8 E(): 0.0023; 26.7% identity in 116 aa overlap"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAB76871.1"

/db_xref="GI:7209209"

/translation="MTAEAPTALSHGRILVPVTHFFDDLDALGLLHNRYPVWVERA WELWAGHGVREGEDEAGDAGNARELTIGVEAPVTRFCAYAVHLWRLTGTGLT YGRFCSADGATTYARGTRVLVRDLARTLRPAWSEAFRTAGRALURPAD"

complement(4404..4407)

4564..6000

/gene="SC01450"

/note="synonym: SCL6.07"

4564..6000

/gene="SC01450"

/note="SCL6.07, probable uracil permease, len: 478 aa;

similar to TR:BAA35783 (EMBL:D90738) Escherichia coli

uracil transport protein PyrP, 442 aa; fasta scores: opt: 908 z-score: 949.9 E(1): 0; 43.1% identity in 429 aa overlap and to SW:PYRP_BACCL (EMBL:X76083) Bacillus caldolyticus uracil permease (uracil transporter) PyrP, 442 aa; fasta scores: opt: 770 z-score: 806.6 E(1): 0; 33.0% identity in 437 aa overlap. Contains Pfam match to entry PF00860 xan_ur_permease, Xanthine/uracil permeases family. Contains also possible hydrophobic membrane spanning regions"

```
/codon_start=1
/transl_table=11
/product="putative uracil permease"
/protein_id="CAB76872.1"
/db_xref="GI:7209210"
/translation="MDLGVYRWKLHGDGKVPACAVRPPDERLSWPTAGLGAQHVVAM
FGASFVAPVLMGLDPLNLAIMMSGVATVIFLLATRGVPSYLGCSLSFVGVAIVRAQG
GTSATQVAVLVVGVAVFLVGLAVRFGARIHAAMPVPTVAVVLMILGFNLAPVTAS
TYWQDMTALFVMTFLVGLVAVCLRGFSRIAFGLGVFVLSWALDRIFGKIHSVD
GSGKPTDWRLLDFSAVGQADWIGLPSFHGSPENWAILVALPVVIALIAENAGHVKA
GEMTGNDLDDKLGTAISADGIGSVLSTAVGPPENTYSENIGVMAATRVYSTAAVAA
AGPALLFGLCPKFGAVVAALPGVGLGGITVILYGMIGLGAQIWINAKVDMNPLNV
PAAAGIIIGVGNVSMETDTFSLSGIALGTLVVITGYHALRAFAPAHKATQPLDDEG
TSYSDAKTGGVGGDGGDGGGAQRARS"
/misc_feature
4654..5793
/genes="SCO1450"
/note="Pfam match to entry PF00860 xan_ur_permease,
```

misc_feature

```
4654..5793
/genes="SCO1450"
/note="Pfam match to entry PF00860 xan_ur_permease,
```

Alignment Scores:

```
Pred. No.: 4,78e-62 Length: 277000
Score: 798.50 Matches: 176
Percent Similarity: 57.52% Conservative: 84
Best Local Similarity: 38.94% Mismatches: 177
Query Match: 33.61% Indels: 15
DB: 1 Gaps: 8
```

US-10-098-602A-2 (1-454) x SCO939109 (1-277000)

```
QY 6 AspGluLeuAsnAenLeuLysLeuGluValValLeuAlaCysVal 25
DB 184054 GAGAACTGACCCCTCGTGGCGCGGTGACATCGACACCTCGCTCGCTTCC 183995
QY 26 AspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeuGlyLeuAsp 45
DB 183994 GACATGAGGGCGCGCTCCAGGGCAAGCGTTCGCGCGCGGTCTTCTCTC 183944
QY 46 GlnLysLysLysSerLysThr-----pheValTyrAlaValThrLeuGlu 61
DB 183943 GATGAGGTCTCTGGAACACCGACCGAGGGCTGCAACTACTCTCTCGCGTGGACGCGGAC 183884
QY 62 GlyLeuAlaGlyGlyTyrGluLysSerValAspThrGlyTyrSerAspCysHis 81
DB 183883 ATGAACACCGTCACCGCTACGCCATGCTCTCTGGGACCGCGCTACGGGCTTCGCG 183824
QY 82 LeuCysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAla 101
DB 183823 ATGCGCGCGGACCGCGCCACCTCGCGCGCGTCCCTGGAACGAGGCGCACCGCGATGGCC 183764
QY 102 IleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValle 121
DB 183763 GTCCGCGGAC---CTCGCTGGGAGGACGGCTCCCGCGGTGCTCGCGCGCGCGCGCGATC 183707
QY 122 LeuMetGlnGlnLeuGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGlu 141
DB 183706 CTGCGCGCGCGCGTGGAGCGGCTCGCGCGGACCGGTACCCGACAGGTGCGCGACCGGAG 183647
QY 142 LeuGluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLysAsn 161
DB 183646 CTGAGGTTCATCGTCTTCGCGGACACCTACGAACACGCGCTGGGACGCGAATACCGCGG 183587
QY 162 LeuLysThrAlaGlnProHisHisGlnTrpMetAsnLysSerAlaSerGlyLeuGlu 181
DB 183586 CTGACCCCGCGGCAACCACTGACACGTCGACTACTCGCTCTCGCGCACCGCGCGCGTGA 183527
```

```
QY 182 ThrPheMetArgSerValArgAsnLysLeuGluAlaGlyLeuMetGluAlaThr 201
DB 183526 CCCCTGCTGGCGCGCTCGCAACGAGATGGCGCGCGCGCTCACCGCTCGAGTCGCGC 183467
QY 202 HisProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAspProLeuThr 221
DB 183466 AAGGCGGAGTGAACCCCGCGCAGCAGATCGCTTCCGCTACGACAGGCGCCCTGGTC 183407
QY 222 MetAlaAspArgHisLeuAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241
DB 183406 ACTGCGGACCGACGCGCTCTACAGACCGCGCGCGCAAGAGATCGCCGCCGAGGGGC 183347
QY 242 MetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisLeu 261
DB 183346 ATGTCCCTCACCTTTCATGCCCAAGTACACGAGCTGGAA---GGCAACTCTCTGCCACATC 183290
QY 262 HisMetSerLeuGlnAspAlaGluThrGluLysAsnAlaPheTyrAspGlnAsnAspGlu 281
DB 183289 CACCTCTCGCTCGCGCAGCGCGAC---GGCGCAACGCGCATGCCGAGGAGGG----- 183239
QY 282 TyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValProGlu 301
DB 183238 ---GGAATGTCGACGCTCATGCGGCATCTCTCGCGGACAGCTGCTGCGCGCTGCCGAG 183182
QY 302 AlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThrPhe 321
DB 183181 TTCTCCCTCTCTACGCGCGCGCACATCACTCTCAAGCGCTTCCAGCGCGTCTCTTC 183122
QY 322 AlaProThrLysCysCysTrpAlaLeuAspAsnArgThrSerAlaPheArgLeuCysAsn 341
DB 183121 GCGCGCGCGCGCTCGCTCGCGCGCGGATTGACCGGATCGAGCGCGCTGGGAATG 182945
QY 342 SerLysSerGluGlyLysValGluLeuValGlyGlyValAspLeuAsnProTyr 361
DB 183064 GGCACGCGCGCTCTCTCGTTTCGAGAACCGCGTCCCGCGGCGGACGTCAACCTCTAC 183005
QY 362 LeuAlaPheSerAlaIleAlaGlyLysSerGlyLeuGluGlyLeuGluLeu 381
DB 183004 CTCGCGGTGGCGCGCTGCTCGCGCGCGGATTGACCGGATCGAGCGCGCTGGGAATG 182945
QY 382 ProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsnSer 401
DB 182944 CCGGAGCGCTCGCGCGCAACGCTACAC---GCCGACTTCGCGCACGCTCCCGCACCC 182888
QY 402 LeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGlyGlu 421
DB 182887 CTGCGGAGCGCGCGAGCTGTGGGAGAACACCGCTCCGCAAGCGCGCTTCGCGGAC 182828
QY 422 LysLeuLeuLeuHisTyrValAsnAlaAlaAsnValGluIleAsnGluPheSerLysGln 441
DB 182827 GAGTCTGCTCGCCACTACCGGAACATGGCCCGCGTCAACTGAGCGCTTCGACGCGCC 182768
QY 442 ValThrAspTrpGluLeuAsnGlnGlyPheAsnArg 453
DB 182767 GTCAACGCTGGGAATGCGCGCTCTTCGAACGC 182732
RESULT 11
AP005047 298550 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 27/30.
DEFINITION AP005047 BA000030
VERSION AP005047.1 GI:29610176
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shimose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
```

avermittilis: deducing the ability of producing secondary metabolites
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
 Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermiltis*
 Nat. Biotechnol. 21 (5), 526-531 (2003)
 22608306
 12692562
 3 (bases 1 to 298550)
 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kishida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.
 Direct Submission
 Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
 (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
 This work was done in collaboration with Haruo Ikeda (*1), Jun Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi Osonoe (*4), Yoshihiro Kishida (*4), Hiseashi Kikuchi (*4), Tadayoshi Shiba (*5), Yoshiyuki Sakaki (*6, *7), Maeahira Hattori (*1, *7) and Satoshi Omura (*1, *3).
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.
 *1 Kitasato Institute for Life Sciences, Kitasato University
 *2 National Institute of Infectious Diseases
 *3 The Kitasato Institute
 *4 National Institute of Technology and Evaluation
 *5 School of Science, Kitasato University
 *6 Institute of Medical Science, University of Tokyo
 *7 RIKEN, Genomic Sciences Center
 Following url is also available.
 http://avermittilis.kitasato-u.ac.jp.
 Location/Qualifiers
 1. .298550
 /organism="Streptomyces avermiltis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227882"
 /note="This strain is also named as strain: ATCC 31267, NCIM 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"
 66. .827
 /note="SAV6513"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC74224.1"
 /db_xref="GI:29610177"
 /translation="MCASENTPRAGGKRAPEGVKHAPEPPAVTVLAVLTIRAGVLQ VLVVERQEPYAGRWALPGGVLPHESAEAAAREAEETGLADLTGLHLEQLRTYSE PDRPRNVSVATALLPDPPEPHGGDAAQWLRYNALPGIAPFDHRLADAHAR VGRKLEYTCLATSCFPPEFTLGELOQVETVWGTALDRPNFRKRLVTPGFEVQIPGA ARLTGGKRGKPAALYRAGGATTLHPPLRLRTSEGRP"
 827. .1849
 /gene="draG2"
 827. .1849
 /gene="draG2"
 /note="SAV6514"
 /codon_start=1
 /transl_table=11
 /product="putative ADP-ribosylglycohydrolase"
 /protein_id="BAC74225.1"
 /db_xref="GI:29610178"
 /translation="MTTTLTKRAATGSLGLGALGDALGFPTFENDVPVSLAKCGPWR

CDS
 1853. .2602
 /note="SAV6515"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC74226.1"
 /db_xref="GI:29610179"
 /translation="MDALDIDILAPVVAQDPVLFATVSGAHLGYGFPSSRSDIDLRG AHLLPAELVGLREPTETSRMMDRGVEMDLVTHDLRKFRVLMRLRNGVYVLEQLLSP LVHTSDAHLAEALAPGVLTRHHHHYRGFATTOMFLFKTGLKPLLTFTFVLLTGL IHLRSGEVOAHLPTLVGQVDAPGWLPELLAAKAREHGAADVQDQARVLDLVERLHGL LDEAGASRLPDAPTAYDALHDLVVRVLESG"
 complement (2528. .3289)
 /note="SAV6516"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC74227.1"
 /db_xref="GI:29610180"
 /translation="MPETDPLGRLVRHHTIYSCVMGSAFGLATDGSOTDRRGVFL APTPLWSPAKPPSHVGEPAEQFSWELERFCBLALRANPNILECLHSLPLVEQADATG REILLAREAFSLQTHETFAVYALSORKLEADVRAHGA PRWKHAMHLRLLTSCRDLLRGSLITDVGDRREPLLVKRGCEVPMPEVSRMTLAAETDEAARTPLPAEPDRRR VDFLTRHASALQAHPDDEVGVVGRGVGGP"
 complement (3311. .3775)
 /note="SAV6517"
 /codon_start=1
 /transl_table=11
 /product="putative iron sulphur protein"
 /protein_id="BAC74228.1"
 /db_xref="GI:29610181"
 /translation="MTIRPTRRTLLGTGAVALLAGSKYKSGDNNSSQSTSGTPSGS PSGTASATAGGAAGPELAKTSDIPVGGGTIFKDEKVVVTQPEQGFKAFAVCTHQ GCTVTVSDGTINCPCHGSKFRITDASVAAGPAPRPLPAEKITVEGNSIRLT"
 complement (3772. .4356)
 /note="SAV6518"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC74229.1"
 /db_xref="GI:29610182"
 /translation="MPVDNNAATQSPFPQPAHRGPGCAARFLVPALVAVAVGLGAY GKVDHPAGTAFNLGAFSGSTGAVKSWLATAFFALVOLVLSALMYGKLPGPSWAPVLH RWSGRVAFVAVPVAHVAVHLYGTYEPRVLWHSLGCTFFGAFSAKMLLLRAERLP GWLLPIVGLVPAVLGTIWLTSALMFFRTFGVTT"
 complement (4395. .5090)
 /note="SAV6519"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC74230.1"
 /db_xref="GI:29610183"
 /translation="MAGNSPRGPVWKRSGPLFTLLAGLLLVFMLSINAMTGTKSTS IYQEEPPAATVSPSPRSTPPSPSPSPADYAGRTDDNSAVAVSRDVGKA IAYFCGRNKESWLKGDVKGDTGMLTKGQAEGLKGEKRI RGTVDLGGTRYGPT ADKALKPSGLYRATASVRGAKIDGGWIVLPGGKQVGLNRDGPQAAPRIDPETGAVT VDGQQLTARPVTP"
 5301. .5663
 /gene="aroH"
 5301. .5663
 /gene="aroH"
 /note="SAV6520"
 /codon_start=1
 /transl_table=11
 /product="putative chorismate mutase"
 /protein_id="BAC74231.1"
 /db_xref="GI:29610184"
 gene
 CDS

gamma-glutamylisopropylamide synthetase; ipuA gene; ipuB gene; ipuC gene; ipuD gene; ipuE gene; ipuF gene; ipuG gene; ipuH gene; putative aldehyde dehydrogenase; putative cytochrome P450; putative ferredoxin; putative permease; putative reductase.

SOURCE ORGANISM

Pseudomonas sp. KIE171
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE AUTHORS

1 de Azevedo Wasch, S.I., van der Ploeg, J.R., Maire, T., Lebreton, A., Kiener, A. and Leisinger, T.

Transformation of isopropylamine to L-alanine by Pseudomonas sp.

strain KIE171 involves N-glutamylated intermediates

Appl. Environ. Microbiol. 68 (5), 2368-2375 (2002)

JOURNAL MEDLINE

21972288
11976110

REFERENCE AUTHORS

2 (bases 1 to 11355)
de Azevedo Wasch, S.I.

Direct Submission

Submitted (03-MAY-2001) de Azevedo Wasch S.I., Microbiology, Swiss

Federal Institute of Technology Zurich, Schmelzbergstrasse 7, 8092

Zurich, SWITZERLAND

FEATURES source

1. .11355
Location/Qualifiers

/organism="Pseudomonas sp. KIE171"

/mol_type="genomic DNA"

/strain="KIE171"

/db_xref="taxon:159091"

1362. .2342

/genes="ipuA"

1362. .2342

/genes="ipuA"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="putative reductase"

/protein_id="CAC81333.1"

/db_xref="GI:15865462"

/db_xref="SPTREMBL:Q936T2"

/translations="MSNAKEYDLIVVAGPIGILYAAAYAGFRLGKLTALFDGLPQVGGQ

VATMYPEKIHVAGFPGIKGDFVKNLLEQSORQDYDLSELIVLSLSLSDDTFSI

TDTRGNHYTAAGVIAAGLCKTPRLPALGVDSFNIHFVPLDSVLGNDVVIAGG

GDSADVAAKAAVTPRAKSVTVIHRARFRAHEASVKDMVSGVVRVAPGVEAAVHEENG

HEFLTEWSGDKKELNFDSPVMAFGHSDLGPMESWGLGIEGRIIPVKNMETNISRV

FAIGDISVEYEGKRLIANGVGEAAIAVNAAVISPELSVLPTHSTNEVK"

2342. .2680

/genes="ipuB"

2342. .2680

/genes="ipuB"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="putative ferredoxin"

/protein_id="CAC81334.1"

/db_xref="GI:15865463"

/db_xref="SPTREMBL:Q936T1"

/translation="MPYVVISDPCIKSRDQACVDVCPVDCIYEAEGLNIQDPECTEGC

ACESVCVFAICTVDAESDDQDLREAREFDIVLPGCDGFIGSPGAQKVRIKSDHP

RVLKIIGVD"

2743. .4122

/genes="ipuC"

2743. .4122

/genes="ipuC"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="gamma-glutamylisopropylamide synthetase"

/protein_id="CAC81335.1"

/db_xref="GI:15865464"

/db_xref="SPTREMBL:Q936T0"

/translations="MSEENKQILKVRDIEKHNIDITIRLGAVIDGVWRGKQVGAEB

FLNKAALDTQISNLIFGWDVADHVLVDGLEFTGWDSDGYDIALIPDLSTLSLVPWQEK

TASVLCDIOHLNGEPLNLSPRNLRLKAIBKAEOLGYKCYAAVEFEFYLNDISIASIA

gene

4194. .5354

/gene="ipuD"

4194. .5354

/gene="ipuD"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="putative cytochrome P450"

/protein_id="CAC81336.1"

/db_xref="GI:15865465"

/db_xref="SPTREMBL:Q936S9"

/translation="MKDVNEVARNDFHGEALDEIFTYSTLRNCGPVGRSNGYGGFW

FLTKGDDIFAAEQDPEAFVSPVMVPSVSEGIQIPDIDPEPTAVRRILLPLFTP

QELKKLEQIRDTARKLAEDFAKEGTGADASVHSRPLPTIIFSLAGVPEKDWPKFD

KWDDIIVYERKEPEVANOAKDVSEYFENLLDNKNGESANLMDYLCKRAXIDGRPL

TRDELLRYCYLLFLAGLDTTANSIRAGLWYLANPEDQOKLNDNELPILPLACEERLRT

LSPVQVMARTCLKQTVIRGQDIKAGERVNLVFGAGNRDEEIFFNPKDIDIEQENRHL

AFGGIHRCLGSLNRLVGIEFLRAVPQKPADPSEKWHGVPKLAF"

5371. .5565

/gene="ipuE"

5371. .5565

/gene="ipuE"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="putative ferredoxin"

/protein_id="CAC81337.1"

/db_xref="GI:15865466"

/db_xref="SPTREMBL:Q936S8"

/translation="MKVTDNNRCQGHARCVYFAPEVFEDDDEGYSRVKAGYETVPVE

LQESVKACANCPALAIKIS"

5589. .6476

/gene="ipuF"

5589. .6476

/gene="ipuF"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="gamma-L-glutamyl-L-1-hydroxyisopropylamide

hydrolase"

/protein_id="CAC81338.1"

/db_xref="GI:15865467"

/db_xref="SPTREMBL:Q936S7"

/translation="MEKRLILICDNTADRASFKKFGVCAPSKEFSEKLLKNYSQIR

TEIAPADPGPLMTPLGAYDGLITGSNSHIYEAQGNLROIIEFAQAFASGTPMFG

VCWGMQLAVAAAGGVLPSRVADSCETPFATGVELTYSGSGHPHHSRTSDFVSP

HSDEVTRLPFGAVVTARNRNFIOAVEIKHGRSTFWGQVTHPELSGDQDQGFURESARS

LVEDGSETLNVHEHAQAQIISMFKAQAISEENLVHFEVDVNTSFEFRPLETLNMLDH

LVIPTAKRFKGGGGLQK"

6533. .7963

/gene="ipuG"

6533. .7963

/gene="ipuG"

/function="isopropylamine degradation"

/codon_start=1

/transl_table=11

/product="putative permease"

/protein_id="CAC81339.1"

/db_xref="GI:15865468"

/db_xref="SPTREMBL:Q936S6"

/translation="MTSTVSAQLKGNLGPIGIALMVATAAPLTVMVGSPMIIGLGN

GVAPMDAVFVGIVMLPSVGVSMKVIENAGAFVAILKGMGVGVGLGAASLAVMS

YTLILALEENYIGVLSGTLSLACISLPWLYTIGIVAFVGLGVNVELSAKVLGV

ALVREISGIVILDLDSVIGSVSGMGTRPFPEISAFILSGPGLGILFAIFGPIGFESTV

YVRESAPERSIPATYIAVIFITLLYFISLWCVVGVGVDDVVGVNNAEGMTFLD

LVAKYLGHVMDLAQVLLITSLFAVVIHINIVARYKYVGLSGCVMLAQAKVHATHS

SPYVASVQTVFVSLLAAATMGDPVPTQIYAWGAAAGTTLGYMIIVLALACFSIICFF

gene

2743. .4122

CDS

2743. .4122

gene

2743. .4122

CDS

2743. .4122

RESULT 14.	
AF440524	128136 bp DNA linear BCT 12-NOV-2002
LOCUS	Pseudomonas aeruginosa strain SG17M genomic sequence, integrated
DEFINITION	Pseudomonas aeruginosa PGI-3 (SG).
ACCESSION	AF440524
VERSION	AF440524.1 GI:24461644
KEYWORDS	
SOURCE	Pseudomonas aeruginosa
ORGANISM	Pseudomonas aeruginosa
	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
	Pseudomonadaceae; Pseudomonas.
REFERENCES	1 (bases 1 to 128136)
AUTHORS	Larbig,K.D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wihlmann, B., Fritz, H. J. and Tumber, B.
TITLE	Genes Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity on a Pseudomonas aeruginosa Clone
JOURNAL	J. Bacteriol. 184 (23), 6665-6680 (2002)
PUBLISHED	12426355

```

REFERENCE 2 (bases 1 to 128136)
AUTHORS Larbig,K.D., Christmann,A., Johann,A., Hartsch,T., Merkl,R.,
        Klockgether,J., Fritz,H.-J. and Tuemmler,B.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2001) Klinische Forschergruppe, Medizinische
        Hochschule Hannover, Carl-Neuberg-Strasse 1, Hannover 30623,
        Germany

FEATURES             Location/Qualifiers
     source            1..128136
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="SG17M"
                        /db_xref="taxon:287"
                        /map="between genes PA2836 and PA2815"
                        /clone="C"
                        /notes="hypervariable genome region"
     misc_feature      1..15865
                        /notes="corresponding genomic region to Pseudomonas
                        aeruginosa PAO1 from gene PA2836 to gene PA2820 with an
                        average nucleotide substitution rate of 0.27%; all
                        annotated genes are conserved and show only a few amino
                        acid substitutions"
     tRNA              15712..15787
                        /product="tRNA-Glu"
     tRNA              15790..15865
                        /product="tRNA-Gly"
     repeat_region     15842..15865
                        /notes="att L site; flanks Pseudomonas aeruginosa SG17M
                        gene island PAGI-3(SG); 24 terminal nucleotides of the 3'
                        end of the Gly-tRNA gene"
                        /rpt_type=direct
                        /rpt_type=direct
                        /rpt_type=direct
     misc_feature      15865..119169
                        /notes="Pseudomonas aeruginosa SG17M gene island
                        PAGI-3(SG), integrated into the first Gly-tRNA gene"
                        /rpt_type=direct
                        /rpt_type=direct
                        /rpt_type=direct
     repeat_region     16091..17500
                        /notes="ORF SG1"
     gene              16091..17500
                        /genes="int"
                        /notes="bacteriophage P4 integrase subfamily; similar to
                        Xylella fastidiosa XF1718 (int) and Pseudomonas aeruginosa
                        C ORF C1"
                        /codon_start=1
                        /evidence=not_experimental
                        /transl_table=11
                        /product="putative integrase"
                        /protein_id="AAN62223.1"
                        /db_xref="GI:24461654"
     misc_feature      20700..24477
                        /translation="MALTDVAVRQARPTGKDYTLDPFDGLSLAVSANGKSGWHFRYYW
                        AKQKMSLGTYPASVLRARALRDEARALAKGINPKVDKXKLRVRLATENSFKA
                        VYLQMLHLEKRGQSTLSQIRIFDKDVLPLPGTGTIFDIRSNLLDLVLAIEQ
                        RGAFTAEKVRWTFRPFAMVKAAGLVGNPASDLVVAAPKPPVAHNPFLRLHEL
                        ELRLKRGVNTTGLGIRLLLTGVRTELELATPDQFDLGRGLWIIIPPEVKQLQ
                        DGNRKGKRPQDIPPIVLSVQAIIVRYLLEQVPAQRHLLAHRGDLKXRISENTL
                        NAALRMGYEGLTGHTGTSTGNEIGYPKIWDVQDLSHADPNKVSATYNADIY
                        EPRKMQWADRLDLLEQGDVKAASLHLAIHINGTQVQNESEGFLAIEGSAGTSVGN
                        SAORNLGGEATWKTGQMA"
                        /complement(17774..18835)
                        /gene="hemE"
                        /notes="ORF SG2"
     gene              22070..22076
                        /complement(17774..18835)
                        /gene="hemB"
                        /notes="ORF SG8"
     CDS               22077..22104
                        /codon_start=1
                        /evidence=not_experimental
                        /transl_table=11
                        /product="putative uroporphyrinogen decarboxylase"
                        /protein_id="AAN62224.1"
                        /db_xref="GI:24461655"
                        /translation="MAELKNDRLRLALLREPVDVTPVMMRQAGRYLPEYRHRARAG
                        DFLSLCKNRELACEVTLPQERYDLDAAILFSDILTIPDAMGLGLYFEEGEGPRFKKT

IRTEVEADLTVPDAASDLGYVLAARKEIRSAAGRVPLIGFSGSPWTLATVYVGGGS
KDFPHIKOMMYATPEVLHALLDLKLAQTVTGLNEQIRSGVQAVQIFDTWGVLSGPL
VLEPSLRVYKQIVDGLIRESGRVPVILFTKNGGQWLEAMAEACADALGLDWTNID
DARRIGRVALQNLDPGALYAPTKRAEVADILARFGSSGSHVFNLGHGHIQYVD
PEVVKVFDVAHELKSPYH"
/gene="ORF SG3"
19225..19680
/gene="ORF SG3"
19225..19680
/gene="ORF SG3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN62225.1"
/db_xref="GI:24461656"
/translation="MNQREQSDNQVHLRLRWFLALTWASIALAVLGAALPGLPPTT
VFLMLIAAWSASRCSQPLRWLEQHPFGDRLRNWQGVGVIDRGSKWMAASAGMLVSMAI
VLLSNQPLLLVPIIATITGAVIVMSRPEOSPMNSQLMQFDKNTY"
20010..20996
/gene="ORF SG4"
20010..20996
/gene="ORF SG4"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein with methyl-accepting
domain of chemotaxis transducer"
/protein_id="AAN62226.1"
/db_xref="GI:24461657"
/translation="MLFGKVKLSDEITNYLSALMSGQIOFVRVDSLINQHPDLVE
LEDIQKWIDQSGIAPSEPESTDELAANLRSCILNAQAEISYLRHELESSRLLDE
ASHELESYRAEIIWDLTKOTLTCGWEIIVIDGIDINPANKLRKSKOFRELIGVSTE
PELDGWSVSVIHPDDIKRVIKALDDYARIGNFDSYVVEYRMRHKIKGDTWYREKG
RGVQPSQLMRIIAGLRDISDEKLAEMHARELENIQATYQISKVVGAIKGLADQT
NMALNAAEARAGVGDFGVKVADEKVLKAGRTREATQKIQEMLSDSKAF"
/complement(21066..21818)
/gene="ORF SG5"
21066..21818
/gene="ORF SG5"
/notes="originally longer protein; the integration of an
IS-element disrupted the CDS (ORFs SG8 and SG5)."
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN62227.1"
/db_xref="GI:24461658"
/translation="MTELRRAINLDELFLVFPQKMGMTTRICGAELARWRHPSHGL
MSPDQIFYAERSGLITRLTYWAIGAALRESYVHSSGEAVPIAVNLSSHDLRDPNLM
AHISSELTWGAHPDWIQFELTESCLMEDIANSMVLKMLRDLGFKLFDIDFGTYSS
LSYLRKLPVNTIKIDQSFVNLGSDSDSAIIVRSTIDLAHNLGLEVVAEGVESKIAMG
MLAENGCEVQGFISKPIPGSDPQDRLSFH"
22070..24477
/notes="insertion sequence of the IS21 family; the
insertion sequence is flanked by inverted repeats and
contains a transposase and a NTP-binding protein; upon
insertion into ORF (SG5 and SG8) a sequence of 7 bp
(CCTAGT) was duplicated
/evidence=not_experimental
22070..24477
/notes="located on the gene island PAGI-3(SG)"
/insertion_seq="IS21 family member"
22070..22076
/notes="flanks insertion sequence"
/rpt_type=direct
22077..22104
/notes="terminal IR of insertion sequence"
/rpt_type=inverted
22083..22137
/notes="repeats L1 and L2 are inverted with respect to R1
and R2 found in the repeat region at the other end of the
insertion sequence"
/rpt_type=direct

```



```

repeat_region 22083..22103
/note="L1"
/rpt_type=direct
repeat_region 22117..22137
/note="L2"
/rpt_type=direct
gene 22183..23688
/gene="ista"
/note="ORF SG6"
22183..23688
/gene="ista"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transposase"
/db_xref="GI:24461659"
/transl_table=11
translation="WISMDIITAEIRRHVSGETISVIARSLNLSRPTVRKHLRSTTA
QVYQROQPAPKLGQFQSTLEAWLNTSEHLPRSORRTARLYEDLQVEGYKGAIDSVQ
RLVKQKALKTRPGAAQAFIPLLPAPGEACQFDWSHQAEIAGVMQTIKVAQFRLCHS
RMFVAYPRETOEMVLDHNRAFAFGVGPQVVIYDNLKTAVDAILVGDRIENRRF
LALANHYLFEPVACTPAAGKEGVQVGNREWLFTPLARFASADLHNLATRCQ
ELAQKPTSRSGIAECFVOEQAHLRIDAPFDGYVEOMSVSSTCLVRVDRNOYSV
PAQAGKVTSRCTADEIRIVADDOLILARHRRGRQOLVDPHLYLAVLDKPGALR
NGAPFVWDLPEPIKQVREYLLKQSRGDRFVLLILLARDVLEALQVACBELAESGV
INGSHVWNEURLRVSARPVELTLPALQRIEPLADQRYEQLRGTOHAY"
gene 23678..24421
/gene="istaB"
/note="ORF SG7"
23678..24421
/gene="istaB"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative NTP-binding protein of insertion
sequence"
/db_xref="GI:24461659"

```

Alignment Scores:

```

Pred. No.: 2e-43 Length: 128136
Score: 595.00 Matches: 154
Percent Similarity: 50.98% Conservative: 80
Best Local Similarity: 33.55% Mismatches: 191
Query Match: 25.04% Indels: 34
DB: 1 Gaps: 9
US-10-098-602A-2 (1-454) x AF440524 (1-128136)
Qy 7 GluLeuAsnLeuIleArgAsnGlyIleAspThrValValLeuAlaCysValasp 26
Db 32888 AACTCCAGGAGTTTCATTGAGAAACACAAATACGACACGGTTAGCGTGGTTCAT 32947
Qy 27 MetGlnGlyArgLeuMetGlyIleArgLeuThrGlyArgHisPheLeu-----Gly 43
Db 32948 ATAGATGAGTCTCGCGGGAACAGATCGGAGCAGAGTACTTTCGAAACAAAGCGCT 33007
Qy 44 LeuAspGlnIleIleSerIleSerIleValThrAlaValThr-----Ile 60
Db 33008 CTTCATGGAACCTCAGATCTCGAATCTTGTTCGATGGGATGTCGCGATCACCCTGTT 33067
Qy 61 GluGlyIleAlaGlyGlyGlyThrGluIleSerValAspThrGlyIleSerAspCys 80
Db 33068 GATGGATGGCGTTTACTGTTGG-----GATTCGCGGTATCCGATATT 33112
Qy 81 HisLeuCyAlaAspLeuAsnSerLeuHisLeuLeuProTTPSerGluGlyAlaValLeu 100
Db 33113 GCTCTAATCCCTGATCTTTGACTCTCTCGCTGTCCTTGGCAGGCAAACTGCATCC 33172
Qy 101 AlaIleSerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgVal 120
Db 33173 GTCCTTTCGATATTCAGCAT--CTCAATGGCCAGTCGCTGACCTTTTCGCCAGGAAT 33229
Qy 121 IleLeuMetGlnGlnIleGluArgLeuAlaAsnLeuIleGlyLeuPheAlaSer 140

```

```

Db 33230 CTCCTGCGCAAGCAACCGAAAAAGAGAGAGAGCTTGCTTACAAAGTGCACGCGCTTAC 33289
Qy 141 GluLeuGluPheAsnLeuPheAsnGluThrTyrrysSerAlaSerGlnIleHisTrpLys 160
Db 33290 GAGTTTGAGTCTACCTTCTGACAGACTCTATTGCGCAGTATTGCGGAGATCAGTGGCGC 33349
Qy 161 AsnLeuLysThrAlaGlnPro-----HisHisGlnTrpMet 172
Db 33350 AGCATTAAATCCAGTGGAAAGAGAGTGGCCACTGTCTACAGCATGTCTCCATCAC----- 33400
Qy 173 AsnIleSerAlaSerSerGlyIleGluThrPheMetArgSerValArgAsnLysLeuGlu 192
Db 33401 -----TCTAGTTCTTCGGACATC-----ATAGTGAAGTTCCGACGTACATCGCT 33445
Qy 193 GluAlaGlyIleLeuMetGluAlaThrHisProGluPheLeuProSerGlnHisGluLeu 212
Db 33446 GATCCCGGGATCGTACTAGAGGCGACGACAGTGAACATGCGCCCGGCGGAGTACAGATC 33505
Qy 213 AsnPheValProAlaAspProLeuThrMetAlaAspArgHisIleIleAlaLysHisGly 232
Db 33506 AACATTAAATGATGACGATGCGCTCAAGACGCGGATGACGCAATTTTGTCAAAAACGCG 33565
Qy 233 ValArgGluMetAlaGluGlnSerGlyMetValAlaThrPheMetAlaLysLeuSerSer 252
Db 33566 ATCAAGAGAGATCGCTGCAAAAGCATGGCATGACGCGACCTTTCATGGCGAAGCCCAATGCA 33625
Qy 253 ThrAlaLeuGlyAsnAlaCysHisIleHisMetSerLeuGlnAspAlaGluThrGluLys 272
Db 33626 GAGTGTGCGGTTTCGTCGGTATGTCACATATGTCGTGAACTTAATTAACAGGTGCC 33685
Qy 273 AsnAlaPheTyrrysGlnAsnAspGluTyrrysGlyMetSerThrLeuAlaAspAsnTrpIle 292
Db 33686 CCAGCATTTGCCAATCCGAGATCCTGCTGCGTGTGTCGAGGTCGGTATTAATCTCTC 33745
Qy 293 AlaGlyLeuLeuLysTyrrysValProGluAlaThrTyrrysPheAlaSerTyrrysAsnSer 312
Db 33746 CGAGCATGTTGAGTTCGACAGAGATTTTTCGCGCATCTATCTGCGCAATATCAACTCC 33805
Qy 313 TyrrysArgLeuGlnProLeuThrPheAlaProThrLysCysTyrrysTrpAlaIleAspAsn 332
Db 33806 TACAGCGGACGCTGCGGATCCTGCGGCGGTGGTAAATTCGTCTTGGGATTCGACAAAC 33865
Qy 333 ArgThrSerAlaPheArgLeuCysAsnSerLysSerGluGlyIleAsnValGluLeuArg 352
Db 33866 CGCACTGTGTCGACCGTGCATCAGCATCAGCGCGCGCGCTCGGTGCGAAGCCGG 33925
Qy 353 IleGlyGlyAlaAspLeuAsnProTyrrysLeuAlaPheSerAlaIleIleAlaGlyIle 372
Db 33926 ATCCCTGTCGCGATACCAACCTTATCTGTTATTTCGCGCAGCTCTTGTTCGGTCTC 33985
Qy 373 SerGlyIleGluGluLysLeuGluLeuProProAlaSerGlyAsnValTyrrysAsp 392
Db 33986 TATGGCATCAGAACCAAACTCAAGCCGAAAGACCCCATCTCTCGGTAACGCTTAC----- 34039
Qy 393 LysGluLeuProGluPheProAsn-----SerLeuGlnAsnAlaThrHisLeu 408
Db 34040 AAGATATCGCGGAGCTCGCACGCTCGCGCGCTCATTTGGAGAGGCTACCAATATC 34099
Qy 409 LeuLysGluSerLysMetLeuAsnLysThrPheGlyGluLysLeuIleLeuHisTyrrysVal 428
Db 34100 TTCGCGCAGAGTGAATGGCAGCAGTGTATTCCCTAAAGAGTTCGTTGAGCATTTATCC 34159
Qy 429 AsnAlaAlaAsnValGluIleAsnGluPheSerLysGlnValThrAspTrpGluLeu 447
Db 34160 CAGATGAAGTTTGGGAGATCAAGCGGCTAATGTTTCGTGAATAATTTGGGAATTG 34216

```

RESULT 15

```

AP005021 300550 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section
DEFINITION 1/30.
ACCESSION AP005021 BA0000030

```


CWLSRSSWLFVRSSTKSRVCPVSRASSCCSPWRFREHRTTPDAAVAPWAAQOPPY
 LAHQFSLPELLELVYGLQOVDPLRIFFPAQVRRVRLDAETGLIDGIRDQ
 QMCPRSTVAVLRMLGRVTRAVRGAHEVYTGAPADDDVLDRALQSRSTPAGIRPK
 TVDLRTIQPVLRLGLLRTWTQORPGADEFARTLGVVELASRALAQRPGTDPAQLRY
 DDVAVDAIFALKRQGEPMYRMSIASHFFNALIDYGRSNAADDLSAARVDRPA
 LHRPEQANEDEIGKALPEPVIKQLDLHLGLGRAGORTLAPDDQUMYITLYI
 LRDTGQRPVLSLPDCLRENDQISLIWNHKKARRRRRLPITTSQAQIETWQ
 RREGQASLPPTGADYLFPAFLTHLATPYLHTSYLGELTRDWVDSIPLHGEGETDQ
 NPLPFRSLIYAYAFRHSYQORHADAGTPLDLVRLMDHKSVSITQRYTVLSLKRKR
 AVTKLAHVHDHGHPPSPSETAYELRSVAVPYGCTEPTSPNVKAAGSSCPTRFOCAGC
 GFYRDPDLYPAISQHINELRADRETAQADAAEFVITAFQAITYAEQVTDNRRL
 AALFVAER"

CDS

complement (9886..10635)

/note="SAV8"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67717.1"
 /db_xref="GI:29603644"
 /translation="MSSISSALSRLPAPRAENAGLDGGWAAWLRHLDPWRTSEW
 RQDCWLFVTSVHEPRSSVALCTEACDTVSPANIFCPCKEOKRSLPLPAEAPARAF
 VVNRNVAFGVGPVPCFTKDGQRCVRPHCKELCATHYTKWTKTKTAGWEDTA
 VPYADTPACAPAPCLPGLXGRGLCRHQAORWMAFLIRLSLAVRISCLASSSVRS
 VAVLVRAGEGVVGCWRCCAAAMCWRMPSA"

CDS

complement (10632..10805)

/note="SAV9"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67718.1"
 /db_xref="GI:29603645"
 /translation="MLRHSATRWLRDGVDRVQVRLIGHASPLSMERYHRVNDAEAR
 AAVRVSGLKERR"

CDS

complement (11101..13497)

/note="SAV10"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67719.1"
 /db_xref="GI:29603646"
 /translation="MLRSLACWPSRATDPSEVSLVRPYAGTTTQWQRLRLLEWAL
 TPGILDAVLIESGHAEARPIAVNSDFWSLYLGAETAPAPARLVGLVLRHHLA
 RARADSDGPFESGHLSTHSDQADTVLSRTAEAPFAYVSQVLPFVIDVATAGTGT
 DAYAPSGRWHLRVGGHVSQALTLTALRSIAASPAADALQOLTPSPVQELR
 FLACRVVPAWQADEALWILNDRNLRLWLSARWASRELIEATTPHGSDLELRL
 TAMLIDYVPAWEHROKSHAWGWSOYELLSAICPSRRSDAGCRLAEWERKEPQGV
 PSPPTIQTGVGSPISDHAARHMTNQWHRALNMNYPQARERFPQGVHLELAQTL
 GSRAEQPERFTDFAFTLGPDPATYLCAIVGAVTPHLDDRWEOALHQLHQLGPA
 APTICRALQSAQNFTAAISLPALDSYTDHPHQDIRDADAEGTRTDLTAGMNATRG
 QAALTVAALLFHGSEHLHALTPLVRLANDSVLAVRVCAQAVALAMKHPDQIALDTT

Alignment Scores:

Pred. No.: 4,4e-39 Length: 300550
 Score: 554.00 Matches: 146
 Percent Similarity: 49.67% Conservative: 79
 Best Local Similarity: 32.23% Mismatches: 219
 Query Match: 23.32% Indels: 10
 DB: 1 Gaps: 6

US-10-098-602a-2 (1-454) x AP005021 (1-300550)

Qy 3 IleThrTyrAspGluLeuAsnAsnLeulleAArgAsnGlyLysIleAspThrValValLeu 22
 Db 227871 GTGACACTCGAAGAGCTCGGCCACCTCGTTCAAGCGCGTGAGATCGACACGCTGATGGTC 227930
 Qy 23 AlaCysValAspMetGlnGlyArgLeuMetGlyLysArgLeuThrGlyArgHisPheLeu 42
 Db 227931 GCCGTCGCGACCTGAGGGCGCGCTGATGGCAAGAACCTCGACGCCACCACTTCTCTG 227990
 Qy 43 GlyLeuAspGlnLysLysIleSerIleSerThrPheValTyrAlaValThrIleGluGly 62
 Db 227991 ACCAAGCTGCCCGCGGACCGAGATGTGCGCCTACATCTCGCGCAGACGCTGGACATG 228050

Qy 63 IleAlaGlyGlyGlyTyrGluIleSerSerValAspThrGlyTyrSerAspCysHisLeu 82
 Db 228051 AACCGCTGGACGGGTTTGAACCTCGCGGTGGCACACCGATACGGAGACCTCGGTGTC 228110
 Qy 83 CysAlaAspLeuAsnSerLeuHisLeuLeuProTrpSerGluGlyAlaValLeuAlaLeu 102
 Db 228111 GTCCCCACCTGGGCACCATCCGCGCTGAACTACCTGCGCAGCACGGTCTCTGTACAC 228170
 Qy 103 SerAsnProHisAsnPheValThrSerGluProLeuPheCysSerProArgValIleLeu 122
 Db 228171 GGCACCGCGTACAC---CCGACGGCGCGCCCTGCGCTGCGCACACGGAGATCTCT 228227
 Qy 123 MetGlnGlnIleGluArgLeuAlaAsnLeuLysLeuLysGlyLeuPheAlaSerGluLeu 142
 Db 228228 CGCAGCAACTCGCGGCGCTCCACGGCTCGCAGGCACGCGTCGCGCATCGAGAGC 228287
 Qy 143 GluPheAsnLeuPheAsnGluThrTyrLysSerAlaSerGlnLysHisTrpLysAsnLeu 162
 Db 228288 GAATTCGTCTCTTCAAAAGGCACCGAAGCACAGCTGCGCGCATCCGGGTTCAGGAACCG 228347
 Qy 163 LysThrAlaGlnProHisGlnTrpMetAsnIleSerAlaSerSerGlyIleGluThr 182
 Db 228348 ACACCGTGTCCCGACCAACCTCGACTACGCGTGGACCGCCCGCCGCTGACCGC 228407
 Qy 183 PheMetArgSerValArgAsnLysLeuGluGluAlaGlyIleLeuMetGluAlaThrHis 202
 Db 228408 TTCTTCACGACCTGCACACGCGCTGCACGAGCAGCGGCACCAATCGAGGCCATCAA 228467
 Qy 203 ProGluPheLeuProSerGlnHisGluLeuAsnPheValProAlaAsp---ProLeuThr 221
 Db 228468 ACGGAAGGAGCCCGACGCGCAGATCGAAGTGGC-CTGCGCCCTACGCGGACACCCCGTGGC 228526
 Qy 222 MetAlaAspArgHisIleAlaLysHisGlyValArgGluMetAlaGluGlnSerGly 241
 Db 228527 GCATGCGACACTACACCTCCACAAACACGCGCTCGTCATCTCGCGGACCAACCGC 228586
 Qy 242 MetValAlaThrPheMetAlaLysLeuSerSerThrAlaLeuGlyAsnAlaCysHisIle 261
 Db 228587 CTCACACCCAGTTTCATGGCC---GCCCGCAGACCGCGCAGCATCGCGGCTCCACCTG 228643
 Qy 262 HisMetSerLeu---GlnAspAlaGluThrGluLysAsnAlaPheTyrAspGluAsnAsp 280
 Db 228644 CACATCTCCCTGTGGCGCGACGACGAG------CGGCGCTTTCGCGCATCGTCCA 228691
 Qy 281 GluTyrGlyMetSerThrLeuAlaArgAsnTrpIleAlaGlyLeuLeuLysTyrValPro 300
 Db 228692 CGCGAGAGCTGCCCGAGCTACTCCACACTTCCATCCCGGCTGTATCGAGCATCGCG 228751
 Qy 301 GluAlaThrTyrPhePheAlaSerTyrIleAsnSerTyrLysArgLeuGlnProLeuThr 320
 Db 228752 CAGCTGGTTCGCTGTACGCGACCCACCCCACTCTTCAAAACGATATGTCGCGCATCTC 228811
 Qy 321 PheAlaProThrLysCysCysTyrAlaIleAspAsnArgThrSerAlaPheArgLeuCys 340
 Db 228812 TTGCGCCCAACCAACTACACTCGGGACGCGCAACACCGCCTCGCGCATCTCGCATACA 228871
 Qy 341 AsnSerLysSerGluGlyIleAsnValGluLeuArgIleGlyGlyAlaAspLeuAsnPro 360
 Db 228872 GGGCCACCGCC---TGCACACACTCGGAAGTACGCTCGCGGAGCGCGCAACCC 228928
 Qy 361 TyrLeuAlaPheSerAlaIleAlaAlaGlyIleSerGlyIleGluGluLysLeuGlu 380
 Db 228929 TACTCGCACTCGCGGACCATCGCGCCATCGTCTACGCGCTCAACACCATCCCCAAG 228988
 Qy 381 LeuProProAlaSerGlyAsnValTyrAsnAspLysGluLeuProGluPheProAsn 400
 Db 228989 CTCCCCGAGCCCTGCACCGCGACGCTACACGCTCCGACCGCTCCCGCTCCCGCCACGC 229048
 Qy 401 SerLeuGlnAsnAlaThrHisLeuLeuLysGluSerLysMetLeuAsnLysThrPheGly 420
 Db 229049 AACCTCGAAGAGCCCTCACCGACTTCCAAAGAGAGCATCGCATCGCGCATTCAGC 229108
 Qy 421 GluLysLeuIleLeuHisTyrValAsnAlaAlaValGluIleAsnGluPheSerLys 440

Db	229109	ACACCCGTAGTGCAGCACTACGCCCGCGCTCCGAGGTTGAGATCGAGGCACAGCACGGC	229168
		
Qy	441	GlnValThrAepTprGluLeuAsnGlnGlyPheAsnArg	453
Db	229169	CAGGTCAAGACATCGACGCGAAGCGGGGGTTCCTCCGT	229207

Search completed: December 18, 2003, 00:54:26
Job time : 6869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:42:10 ; Search time 19 Seconds
(without alignments)
2297.925 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MITYDELNNLRNGKIDTV.....INEFSQVTDWELNQGFNRY 454
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.5	39.7	454	2 AF2839	glutamine syntheta
2	942.5	39.7	454	2 H97616	probable glnA4 pro
3	927	39.0	434	2 A13578	glutamate-ammonia
4	799.5	33.6	457	2 F70885	probable glnA4 pro
5	798.5	33.6	462	2 T36948	glutamine syntheta
6	510.5	21.5	455	2 C84133	glutamine syntheta
7	480	20.2	439	2 B72313	conserved hypothet
8	469	19.7	413	2 C83225	glutamate-ammonia
9	466	19.6	443	1 AJCLOA	glutamate-ammonia
10	465	19.6	476	2 AB3575	glutamine synthase
11	464.5	19.5	449	2 G75517	glutamate-ammonia
12	464	19.5	446	2 H86899	glutamate-ammonia
13	459	19.3	458	2 B83609	probable glutamine
14	458	19.3	458	2 A83391	probable glutamine
15	450	18.9	452	2 D83609	probable glutamine
16	443	18.6	446	2 A43955	glutamate-ammonia
17	442.5	18.6	444	1 AJSBSQS	glutamate-ammonia
18	434.5	18.3	429	2 D90428	hypothetical prote
19	432	18.2	448	2 C95058	glutamine syntheta
20	432	18.2	448	2 D97927	glutamate-ammonia
21	430	18.1	449	2 H83944	glutamine syntheta
22	424.5	17.9	446	2 T51803	glutamate-ammonia
23	423	17.8	446	2 H89905	glutamate-ammonia
24	421	17.7	464	2 AE2650	glutamine syntheta
25	421	17.7	479	2 D97432	probable glutamine
26	412	17.3	445	2 A48947	glutamate-ammonia
27	411.5	17.3	444	1 AJSBSQU	glutamate-ammonia
28	411	17.3	444	2 AH1599	glutamine syntheta
29	410	17.3	444	2 AC1237	glutamine syntheta

ALIGNMENTS

RESULT 1

AF2839
glutamine synthetase glnA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF2839
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, i Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2839
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43132.1; PID:gi7740606; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: glnA
A:Map position: circular chromosome
C:Superfamily: glutamate-ammonia ligase
Query Match 39.7%; Score 942.5; DB 2; Length 454;
Best Local Similarity 43.3%; Pred. No. 5.2e-69;
Matches 195; Conservative 89; Mismatches 165; Indels 1; Gaps 1;
QY 4 TYDELNNLRNGKIDTVVLACVDMOGRMLGKRLTGRHFLGLDQKKISISTFFVAVTIEGI 63
Db 5 TFDALKMDVAEGRIDTVLACLVDMOGRMLGKRLTGRHFLGLDQKKISISTFFVAVTIEGI 64
QY 64 AGCGYEIISVDYGDCHLCADNLHLLPNSSEGAVLAISNPHNFVTSPLFCSPRVLIM 123
Db 65 TVPGYKSSSWEXGVDYTLKPDLSLRKVPWLEGTAFLVDLDHHTHEEVPHSPRALLK 124
QY 124 QOIERLANLKLGLFASLEFNLFNETHYKASQKHWKLTQAPHHQWNISASSGIETFF 183
Db 125 KQVAREAMGLKAYMATELEFFLPDQTFDAARASGYKDLNLASGYNEDYHIQTITKEEDV 184
QY 184 MESVANKLEEAGILMEATHPEPLPSOHEINFPVADPLTMADRHITAKHGVREMAESGMV 243
Db 185 MEALRKGGAGIPVENSKEASPGQAEINVRVYAEALTWADRHAIKNATKRIAMSKGA 244
QY 244 ATFMAKLSSTALGNACHIMSLQDAETKNAFYDQNDYGMSTLARNWLAGLLKVVPEAT 303
Db 245 VTFLAKWYNNAAGSSSHIHQSLWLD-GKPAFLDKDGEHGMDSVMHYVAGLAHSDIT 303
QY 304 YPFASYNSYKRLQPLTFAPTKCCWAINRISAFILCNKSKSEGINVEIURIIGADINPYLA 363
Db 304 YFLAPYNSYKRFMAGTFAPTAKIWSLDNRITAGYRLCGAETKIRIECRVGSIDINPYLA 363

DB 78 MNTVDGYAMSSWDURGYGDFAMRADPATLRRLPWNEGTAMAVAD-LAWEDGSPVLAAPRQI 136

283

DB 78 MNTVDGYAMSSWDURGYGDFAMRADPATLRRLPWNEGTAMAVAD-LAWEDGSPVLAAPRQI 136

Matches 133; Conservative 71; Mismatches 208; Indels 52; Gaps 10;

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: A83391

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-458 <STO>

A;Cross-references: GB:AE004630; GB:AE004091; NID:g9948041; PIDN:AAG05428.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA2040

C;Superfamily: glutamate-ammonia ligase

Query Match 19.3%; Score 458; DB 2; Length 458;

Best Local Similarity 28.2%; Pred. No. 2.1e-29;

Matches 125; Conservative 85; Mismatches 211; Indels 22; Gaps 10;

Qy 16 KIDTWLACVDMQGRMGRLTGRHFLGLDQKKISISTFVAVTIEG----IAGGGYEIS 71

Db 22 EVQFVDLLISDMGVGVRGKRIENSLPKVPEKGINLPASLPALDITSTVESTGLGIDIG 81

Qy 72 SVDTGYSDCHLCADL-NSLHLLPWSGAVLAISNPHNFVTSEPLFCSPRVILMQOIERLA 130

Db 82 DAD-----RICYPITPGTSLMEPQKRPQAQLLMTMEHELEGEPPFADPREVLQVVARFT 135

Qy 131 NLKLGIFASELEFNLFN-ETYSASQKHWNKLTQAHPHOMNISASSGIETFMRSVRN 189

Db 136 EMELTIVAFALEFYLIDQENVNGRPQPPRSPISGRP--QSVQVYSIDDLDEYVEGLQD 193

Qy 190 KLEEA---GILMEATHPEFLPSQHELNF-VPADPLTMADRHIIAKHGVREMAEQSGMVAT 245

Db 194 IIDGARAQIPADAIVAESAPAEVNLHVADPMKACDVAVLKRLIKNIAYDHEMDTT 253

Qy 246 FMAKLSLTALGNACHIHMSLQDAETEKNAFYDQNDYGMSTLARNIAGLLKYVPEATYF 305

Db 254 FMAKPYEQAGNGLHVHISLLD-KHGNNIFTSDEPE--QNAALRAHIGGVLETLPASMAF 310

Qy 306 FASVINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNSKSEGINVELRIGGADLNPYLAFS 365

Db 311 LCPNVNSYRRFGSGFYVPNAPSNGLDNRITVALRYPTGSPDAVRLHVRVAGADANPYLLA 370

Qy 366 AIIAAGISGIEEKLPPPPASGVNVDKELPEPPNSLQNAATHLLKESKMLNKTGKLLIL 425

Db 371 SVLAGVHGLTNKVEGPAIEGNSYEQLE-PSLPNNLRDALRELDSDSEILAKYIDPKYID 429

Qy 426 HYVNAANVEINEFSKQVTDWELN 448

Db 430 IFVACKESLEEFYSISDLEYN 452

RESULT 15

D83609

probable glutamine synthetase PA0298 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001

C;Accession: D83609

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: D83609

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-452 <STO>

A;Cross-references: GB:AE004467; GB:AE004091; NID:g9946133; PIDN:AAG03687.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0298

C;Superfamily: glutamate-ammonia ligase

Query Match

Best Local Similarity 18.9%; Score 450; DB 2; Length 452;

Matches 130; Conservative 80; Mismatches 222; Indels 26; Gaps 7;

Qy 1 MTITYDELNNLIRNGKIDTIVVLACVDMQGRMGRLTGRHFLGLDQKKISISTFVAVTI 60

Db 1 MTTKLDQLTSMWKERKITEVECLISDLTGIARGKISPTNKFIAEKGMRLPESVLLQITVTG 60

Qy 61 EGIAGGGYEISSTVDTGYSDCHLCADLNSLHLLPWS-EGAVLAISNPHNFVTSEPLFCSPR 119

Db 61 DYVEDDIY-YDLLDPADIDMVCRPDENAVFLVPAIEPTAMVIHDTFDKL-GNPIELSPR 118

Qy 120 VILMQOIERLANLKLGLPASELEFNLFN-----YKSASQKHWNKLTQAHP 168

Db 119 NILKRVLKMYADKGNRPPIVAPENEFYLTCKSDDDPYLPQAPVGRSGRQ-----ETG 169

Qy 169 HOMWNISASSGIETFMRSVRNKLKEEAGILMEATHPEFLPSQHELNFVPADPLTMADRHII 228

Db 170 RQSFSDAANEFDPFEDMYDWCEAQCGLDLDTLIHEEGTAQMEINFRHGDALDLADQILV 229

Qy 229 AKHGVREMAEQSGMVATFMAKLSLTALGNACHIHMSLQDAETEKNAFYDQNDYGMSTLA 288

Db 230 FKRTMREAAALKHNVAATFMAKPMTGEPGSAMHLHQISIVDVTKGNIF--SNADGTMSELF 287

Qy 289 RNWIAGLLYVPEATYFFASVINSYKRLQPLTFAPTCKCWAIDNRTSAFRLCNSKSEGIN 348

Db 288 LHHIGGLQKFIPEVLPLFAPNVNSFRFLPDTAPNVNVEGEENRTVGLRVDPSSPENRR 347

Qy 349 VELRIGGADLNPYLAFSAIIAAGISGIEEKLPPPPASGVNVDKELPEPPNSLQNAATHL 408

Db 348 VENRLAGADANPYLAAASLLCGYIGWVEGIKPSAQVKGGRYERNL-RLPLTIEAALER 406

Qy 409 LKESKMLNKTGKLLILHYVNAANVEINEFSKQVTDWE 446

Db 407 MENCKPLEQYLGSKFISGYVAVKRAHENPKRVISSWE 444

Search completed: December 17, 2003, 22:44:47

Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2003, 22:40:45 ; Search time 38 Seconds
(without alignments)
3083.050 Million cell updates/sec

Title: US-10-098-602A-2
Perfect score: 2376
Sequence: 1 MTTYDELNLRNGKIDTV.....INBFSQVTDWELNQGNRY 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	42.7	465	16	Q98H15
2	969.5	40.8	455	16	Q8FVU3
3	963	40.5	454	16	Q92NI9
4	942.5	39.7	454	16	Q8UD13
5	927	39.0	434	16	Q8YCH5
6	799.5	33.6	457	16	Q33342
7	798.5	33.6	462	16	Q88070
8	614.5	25.9	459	2	Q93670
9	595	25.0	460	2	Q8GPY7
10	510.5	21.5	455	16	Q9K665
11	470.5	19.8	427	2	Q9AEL4
12	470.5	19.8	446	16	Q8NNH4
13	470.5	19.8	516	16	Q8NHL7
14	469	19.7	413	16	Q9HYP0
15	465	19.6	476	16	Q8YCK5
16	465	19.6	476	16	Q8FVQ7

17	464.5	19.5	449	16	Q9RX64
18	459	19.3	458	16	Q916J5
19	458	19.3	458	16	Q91275
20	453	19.1	445	16	Q8G4S3
21	453	19.1	461	16	Q8P897
22	451	19.0	448	16	Q8DVU9
23	450	18.9	452	16	Q916J3
24	450	18.9	478	16	Q92RY1
25	449.5	18.9	447	17	Q8PY99
26	449.5	18.9	483	17	Q976N5
27	449	18.9	448	17	Q9H149
28	449	18.9	461	16	Q8FJQ2
29	447.5	18.8	445	16	Q8EQP3
30	447.5	18.8	455	16	Q92MV2
31	446.5	18.8	491	17	Q8TYG1
32	445	18.7	448	16	Q8DXT3
33	445	18.7	451	16	Q8EHF4
34	444.5	18.7	447	17	Q8TID7
35	442	18.6	446	16	Q98EM0
36	441	18.6	448	16	Q99Y41
37	441	18.6	448	16	Q8NZG4
38	440	18.5	448	16	Q8E3F3
39	434.5	18.3	429	17	Q97VQ9
40	434	18.3	462	17	Q977Z8
41	432	18.2	448	16	Q97S86
42	432	18.2	448	16	Q8DQX6
43	430	18.1	449	16	Q9KAC9
44	430	18.1	478	16	Q988Z8
45	425.5	17.9	453	16	Q9RDS6

ALIGNMENTS

RESULT 1

Q98H15 ID Q98H15 PRELIMINARY; PRT; 465 AA.
AC Q98H15;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Glutamine synthetase;
GN MLL3074;
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR ENBL; AP003001; BAB50051.1; -;
DR InterPro; IPR001637; GlnA adenyltn.
DR InterPro; IPR001691; Gln synth.
DR Pfam; PF00120; gln-synt_1;
DR Pfam; PF03951; gln-synt_N; 1;
DR ProDom; PD001057; GlnA adenyltn; 1.
KW Complete proteome.
SQ SEQUENCE 465 AA; 52610 MW; 99A46709D3D184FA CRC64;

Query Match 42.7%; Score 1015; DB 16; Length 465;
Best Local Similarity 45.7%; Pred. No. 1e-75;
Matches 206; Conservative 81; Mismatches 164; Indels 0; Gaps 0;

OY 4 TYDELNLRNGKIDTVTLACVDMQGRMGKRLTGRHFLGLDQKTKISISFVYAVTIEGI 63

QY 121 ILMOQIERLANLKLGLFASLEENLNFTYKSAQKHWNKLTAPQPHQHMNTSASGI 180
 DB 121 ILKQVAREANGFKAYMASSELEFFLDQSDYDARLQGLASGYNEDYHIFQTTKE 180
 QY 181 ETFMRSVRNKLSEAGILMEATHPEFLPSOHLNFPADPLTMADRHIIAKHGVRMAEQS 240
 DB 181 EDVNRARNGLOAGIPVENSKGASAGQEEINRYADALTMADRHAIKNGCKEIAWQR 240
 QY 241 GMVATFWAKLSSTALGNACHTHMSLQDAETKNAFYDNDYGNSTLARNWAGLLKVP 300
 DB 241 GKAITFLAKWNSAAGSSSHIHQSLMSKOGTFLFFDNGQYGNSELMRHRVVAQLAHAS 300
 QY 301 EATVFFASYINSYKRLQPLTAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGADLNP 360
 DB 301 EVTYFLAPYINSYKRFNAGTAPTKAIWSKDNRTAGYRLCGESKAIKRIEYCRVGSGLNP 360
 QY 361 YLAFSAIIAGISGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFF 420
 DB 361 YLAFSAIIAGISGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFF 420
 QY 421 EKLILHYVNAANVEINFSKQVTDWELNQGFR 453
 DB 421 EKVVDHYVHAARWEQYEDRRVTDWEVARGFER 453

RESULT 4
 Q8UDI3
 ID Q8UDI3 PRELIMINARY; PRT; 454 AA.
 AC Q8UDI3;
 DT 01-JUN-2002 (TrEMBLrel. 21; Created)
 DT 01-JUN-2002 (TrEMBLrel. 21; Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23; Last annotation update)
 DE Glutamine synthetase.
 GN GLNA OR ATU2142 OR AGR C 3883.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58.";
 RT Science 294:2317-2323 (2001).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328 (2001).
 DR ENBL; AE009164; AAL43132.1; -
 DR ENBL; AE008130; AAK87889.1; -
 DR InterPro; IPR001637; GlnA_adenyln.
 DR InterPro; IPR001691; GlnA_synth.
 DR Pfam; PF00120; gln-synt; 1.
 DR Pfam; PF03951; gln-synt; 1.
 DR ProDom; PD001057; GlnA_adenyln; 1.
 DR ProDom; PD001057; GlnA_adenyln; 1.

KW Complete proteome.
 SQ SEQUENCE 454 AA; 50703 MW; 559465BB5613E37E CRC64;
 Query Match 39.7%; Score 942.5; DB 16; Length 454;
 Best Local Similarity 43.3%; Pred. No. 1.1e-69; Mismatches 165; Indels 1; Gaps 1;
 Matches 195; Conservative 89;
 QY 4 TYDELNNLRNGKIDTVVLACVDMQGRMLGKRLTGRHFLGLDOKKISISTFVYAVTIEGI 63
 DB 5 TFDALKMDVAEGRIDTVLACVDMQGRMLGKRLTGRHFLGLDOKKISISTFVYAVTIEGI 64
 QY 64 AGGGEYISSVDVTGYSDCHLCADLNSHLPLWSEGAVLAISNPHNFTVSEPLFCSPRVLIM 123
 DB 65 TVPGYKSSSEWEGYDYLTKPDLSTLRKVPWLEGTALVLCVDLDDHHTHEEVPHSPRALLK 124
 QY 124 QQTERLANLKLGLFASLEENLNFTYKSAQKHWNKLTAPQPHQHMNTSASGIETFF 183
 DB 125 QQVAREANGFKAYMASSELEFFLDQSDYDARLQGLASGYNEDYHIFQTTKEEDV 184
 QY 184 MRSVRNKLSEAGILMEATHPEFLPSOHLNFPADPLTMADRHIIAKHGVRMAEQSGMV 243
 DB 185 MRALRKGLQAGIPVENSKGASAGQEEINRYADALTMADRHAIKNGCKEIAWSKGA 244
 QY 244 ATFWAKLSSTALGNACHTHMSLQDAETKNAFYDNDYGNSTLARNWAGLLKVPBEAT 303
 DB 245 VTFLAKWNSAAGSSSHIHQSLMSD-GKPAFLDKGEGHGMDSVMRHRVVAQLAHASDIT 303
 QY 304 YEFASYINSYKRLQPLTAPTKCCWAIDNRTSAFRLCNSKSEGINVELRIGADLNPVLA 363
 DB 304 YFLAPYINSYKRFNAGTAPTKAIWSDNRTAGYRLCGESKAIKRIEYCRVGSGLNPVLA 363
 QY 364 FSAIIAGISGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFFGEKL 423
 DB 364 MALLAAGIDGIEKLELPPASGNVNDKELPEPPNSLONATHLLKESKMLNKTFFGEKL 423
 QY 424 ILHYVNAANVEINFSKQVTDWELNQGFR 453
 DB 424 VDHVHAARWEQYEDRRVTDWEVARGFER 453

RESULT 5
 Q8YCH5
 ID Q8YCH5 PRELIMINARY; PRT; 434 AA.
 AC Q8YCH5;
 DT 01-MAR-2002 (TrEMBLrel. 20; Created)
 DT 01-MAR-2002 (TrEMBLrel. 20; Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22; Last annotation update)
 DE Glutamine synthetase (EC:6.3.1.2).
 GN BMEI10554.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
 RA Selkov E., Elzer P.H., Hagus S., O'Callaghan D., Letesson J.-J.,
 RA Hasekorn R., Kyrides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
 RL ENBL; AE009691; AAL53796.1; -
 DR InterPro; IPR001637; GlnA_adenyln.
 DR InterPro; IPR001691; GlnA_synth.
 DR Pfam; PF00120; gln-synt; 1.
 DR Pfam; PF03951; gln-synt; 1.
 DR ProDom; PD001057; GlnA_adenyln; 1.
 DR ProDom; PD001057; GlnA_adenyln; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 434 AA; 49582 MW; 1114E54225746C7A CRC64;

RA Oliver K., Harris D.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +
 L-GLUTAMINE.
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
 DR EMBL; AL939109; CAA20824.1; -;
 DR HSSP; P06201; ILGR.
 DR InterPro; IPR001637; GlnA_adenyltn.
 DR InterPro; IPR001691; Gln_synth.
 DR Pfam; PF00120; gln-synt_1.
 DR ProDom; PD001057; GlnA_adenyltn; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 462 AA; 50818 MW; 595825AA19EBE700 CRC64;
 Query Match 33.6%; Score 798.5; DB 16; Length 462;
 Best Local Similarity 38.9%; Pred. No. 9.9e-58;
 Matches 176; Conservative 84; Mismatches 177; Indels 15; Gaps 8;
 Qy 6 DELNNLRNGKIDITVVLACVDMQGLMKRLTGRHFLGDLQKISIST- - - - FVYAVTIE 61
 Db 21 EELHALVAAGDIDITVVLAFDPDMQGLQKRFPAARFL- - - DEVLEHGTGECNYLLAVDAD 77
 Qy 62 GIAGGVEISSVDTGYSDCHCADLNSLHLLPWSGAVLAISNPHNFTVSEPLFCSPRVI 121
 Db 78 MNTVDGYAMSSWDGRGDFANRADPATLRLPWNEGTAMAVAD- - - LAWEDGSPVLAAPRQI 136
 Qy 122 LMQOIERLANLKLGLFASELEFNLFNETYKSAQKHWKLNKTAQPHHQNWNISASSGIE 181
 Db 137 LRRQLERLAGHYTAQVGTLEFIVFRDTEHNDANYRGUTPANQNVDSVILGTGRVE 196
 Qy 182 TFMRSVRNKLREAGILMEATHPELPQHELFNFPADPLTMADRHIIAKHGVRMAEQSG 241
 Db 197 PLLRRIRNEMAGLTVESAKGECNPGQHEIAFYDEALVTCQDHVYKTKAKGIAQEG 256
 Qy 242 MVATFMKLSLTALGNACHIMSLQDAETKNAFYQNDYGMSTLARNWTAGLLKYVPE 301
 Db 257 MSLTFMAYKNELE- - GNSCHILSLADAD- - GRNMAEGG- - - GMSDVMRHFILAGQVALRE 311
 Qy 302 ATYFFASVINSYKRLQPLTFAPTKCCWADNRITSAFRLCNKSEGINVELRIGADLNPI 361
 Db 312 FSLLYAPHINSYKRFQPGSFATVAGHNDNRITCALRVV- - - GHGSLRFENKLPQGDVNPY 370
 Qy 362 LAFSAITAAAGISGIEKLELPPASGNVYNDKELPEFPNSIQNATHLLKESKMLNKTGFE 421

Db 371 LAVAGLVAAGLHGIEQRLPEPCPGNAYT-ADFAHVPTTLREAAELWENSTLAKAAGD 429
 Qy 422 KLILHYVNAANVEINEFSKQVTDWELNQGNR 453
 Db 430 EVVAHYRNMARVELDAFADAAVTWELRRSFER 461
 RESULT 8
 Q936T0 PRELIMINARY; PRT; 459 AA.
 AC Q936T0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Gamma-glutamylisopropylamide synthetase (EC 6.3.1.2) (Glutamine
 synthetase) (Glutamate--ammonia ligase).
 GN IPUC.
 OS Pseudomonas sp. KIE171.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=159091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIE171;
 RA de Azevedo Waesch S.I., van der Ploeg J.R., Kiener A., Leisinger T.;
 RT "Degradation of isopropylamine by Pseudomonas species KIE171";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +
 L-GLUTAMINE.
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
 CC EMBL; AJ311159; CAC81335.1; -;
 DR HSSP; P06201; 2GLS.
 DR InterPro; IPR001637; GlnA_adenyltn.
 DR InterPro; IPR001691; Gln_synth.
 DR Pfam; PF00120; gln-synt_1.
 DR Pfam; PF03951; gln-synt_N; 1.
 DR ProDom; PD001057; GlnA_adenyltn; 1.
 KW Ligase.
 SQ SEQUENCE 459 AA; 50586 MW; CBEP4392500A2436 CRC64;
 Query Match 25.9%; Score 614.5; DB 2; Length 459;
 Best Local Similarity 34.8%; Pred. No. 1.9e-42;
 Matches 159; Conservative 73; Mismatches 194; Indels 31; Gaps 10;
 Qy 7 ELNNLRNGKIDITVVLACVDMQGLMKRLTGRHFL- - - GLDQKISISTFVAVTIEGI 63
 Db 11 KVRDFIEKHNDITIRLGAVDIDGVRGKQVGAIEFLNKAALDGTQISNILLFGMDVADHLV 70
 Qy 64 AGGVEISSVDTGYSDCHCADLNSLHLLPWSGAVLAISNPHNFTVSEPLFCSPRVLIM 123
 Db 71 - - DGLFTGWDSGYPDIALIPDLSTLSLVPWQKTASVLCDIQH- - - LANGEPLNLSPRNLR 127
 Qy 124 QOIERLANLKLGLFASELEFNLFNETYKSAQKHWKLNKTAQ- - - - - HHQWNIS 175
 Db 128 KATKAEAQGLYKCYAAVEPEFYLLNDSIASISADQWRSINPVEKSGHCYSMLHH- - - - S 182
 Qy 176 ASSGIETFMRSVRNKLREAGILMEATHPELPQHELFNFPADPLTMADRHIIAKHGVR 235
 Db 183 SSSDI- - - MGEVRKYMRDAGIVLEATNSEHGPQGYEINIKYDYDALKAAADAIFVNGIKE 239
 Qy 236 MASQSGVATFMKLSLTALGNACHIMSLQDAETKNAFYQNDYGMSTLARNWTAG 294
 Db 240 IAAKHGTATFMKPSAEWSGSGGHVMSLSDLAGTP- - - VFAPENPGALSEVYGNFLAG 297
 Qy 295 LLKYVPEATYFFASVINSYKRLQPLTFAPTKCCWADNRITSAFRLCNKSEGINVELRIG 354
 Db 298 MVALAREMSAIIYLPNINSYKRTKAGASWAGGNSGWFNRTVSHRAITSAAGSAARVENRIP 357
 Qy 355 GADLNPIYAFSAIIAGISGIEKLELPPASGNVYNDKELPEFPN- - - - - SLQNATHLLK 410
 Db 358 GADTNPYLVIAASLLSGLYGIENKLPKDPILGNAY- - - KVSPELARPLAASLEEAAGIFR 415

AC Q9AEL4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Glutamine synthetase 2 (EC 6.3.1.2) (Glutamate--ammonia
 DE ligase).
 DE GLNA2.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC13032;
 RA MEDLINE=21338478; PubMed=11445173;
 RX Nolden L., Farwick M., Kraemer R., Burkovski A.;
 RA "Glutamine synthetases of Corynebacterium glutamicum: transcriptional
 RT control and regulation of activity";
 RL FEMS Microbiol. Lett. 201:91-98(2001).
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + PHOSPHATE +
 CC L-GLUTAMINE.
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
 DR EMBL; AJ310086; CAC34378.1; --
 DR HSSP; P06201; ILGR.
 DR InterPro; IPR001637; GlnA_adenyltn.
 DR InterPro; IPR001691; Gln_synth.
 DR Pfam; PF00120; gln-synt; 1.
 DR Pfam; PF03951; gln-synt N; 1.
 DR ProDom; PD001057; GlnA_adenyltn; 1.
 KW Ligase.
 SQ SEQUENCE 427 AA; 48146 MW; C85326A81A2406F6 CRC64;
 Query Match 19.8%; Score 470.5; DB 2; Length 427;
 Best Local Similarity 28.9%; Pred. No. 1.5e-30;
 Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;
 Qy 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFYAVYVITIEGIAGGVEISVD 74
 Db 10 GHLKSVVVAPELESAL-----EGIGFD-----GSAIEGYA----RISEAD 47
 Qy 75 TGYSDCHLCADLNSLHLLPWSG-AVLAINPHNFTV---SEPLFCSPRVILMQOIERLA 130
 Db 48 T-----IARPDPTSTFQVLPLEAGISKLQAARLFCVDVMPDQGPSFSDPRQVLRQVQLAA 102
 Qy 131 NLKLGFLFASLEFLNFTNYKSAQKHWNKLT-----AQPHHWMNISASSGIETFM 184
 Db 103 DEGLTCMISPEIEFYLV-----QSLRTNGLPPVPTDNGGYFDQATFNEAPNFR 150
 Qy 185 RSVRNKLEEAAGILMEATHPELPSQHELNFVPADPLTMADRHIIAKHGVRERMAQSGMVA 244
 Db 151 RNAMVALEELGIPVEFSHHTAPGQOEIDLRHADALTMAADNIMTFRYIMKQVARDQGVGA 210
 Qy 245 TFMAKLSTALGNACHIMSLQDAETKNAPYDQNDYGMSTLARNWIAGLLKVVPEATY 304
 Db 211 SFMPKPFQEHAGSAMTHMSLFEGDT--NAPHDPDDSYMLSKTAKQIAGILHHAPEFTA 268
 Qy 305 FFASYINSYKRLQPLTAPTCCWAIDNRTSAFRLCN---SKSEGINVELRIGGADLNPNY 361
 Db 269 VTNQWVNSYKRIYVGNREAPTAATMGVNSRSALVRVPTVRLNKESSRREVRVLPDTACNPNY 328
 Qy 362 LAFSAIIAAGISGIEEKLPPPPASGNV-----YNDKELPEFPNSLQNAHLL 409
 Db 329 LAFSVMLGAGLKGIGKEGYELDEPAEDDISNLSFRERRAMGYND-----LPSSLDQALRQM 383
 Qy 410 KESKMLNKTGFEKILHYVNAANVEINEFQSKVQTDWELNQGDN 452
 Db 384 EKSELVADILGEHVFEEFLRNKREWRDYOQITPWEELRNLD 426
 RESULT 12
 Q8NNH4 PRELIMINARY; PRT; 446 AA.
 ID Q8NNH4
 AC Q8NNH4;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Glutamine synthetase (EC 6.3.1.2).
 DE CGL2229.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032";
 DR EMBL; AP005281; BAB99622.1; --
 DR InterPro; IPR001637; GlnA_adenyltn.
 DR InterPro; IPR001691; Gln_synth.
 DR Pfam; PF00120; gln-synt; 1.
 DR Pfam; PF03951; gln-synt N; 1.
 DR ProDom; PD001057; GlnA_adenyltn; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 446 AA; 50381 MW; 30E54AC65E3B2CF1 CRC64;
 Query Match 19.8%; Score 470.5; DB 16; Length 446;
 Best Local Similarity 28.9%; Pred. No. 1.6e-30;
 Matches 134; Conservative 84; Mismatches 174; Indels 71; Gaps 12;
 Qy 15 GKIDTVVLACVDMQGRMLGKRLTGRHFLGLDQKKISISTFYAVYVITIEGIAGGVEISVD 74
 Db 29 GHLKSVVVAPELESAL-----EGIGFD-----GSAIEGYA----RISEAD 66
 Qy 75 TGYSDCHLCADLNSLHLLPWSG-AVLAINPHNFTV---SEPLFCSPRVILMQOIERLA 130
 Db 67 T-----IARPDPTSTFQVLPLEAGISKLQAARLFCVDVMPDQGPSFSDPRQVLRQVQLAA 121
 Qy 131 NLKLGFLFASLEFLNFTNYKSAQKHWNKLT-----AQPHHWMNISASSGIETFM 184
 Db 122 DEGLTCMISPEIEFYLV-----QSLRTNGLPPVPTDNGGYFDQATFNEAPNFR 169
 Qy 185 RSVRNKLEEAAGILMEATHPELPSQHELNFVPADPLTMADRHIIAKHGVRERMAQSGMVA 244
 Db 170 RNAMVALEELGIPVEFSHHTAPGQOEIDLRHADALTMAADNIMTFRYIMKQVARDQGVGA 229
 Qy 245 TFMAKLSTALGNACHIMSLQDAETKNAPYDQNDYGMSTLARNWIAGLLKVVPEATY 304
 Db 230 SFMPKPFQEHAGSAMTHMSLFEGDT--NAPHDPDDSYMLSKTAKQIAGILHHAPEFTA 287
 Qy 305 FFASYINSYKRLQPLTAPTCCWAIDNRTSAFRLCN---SKSEGINVELRIGGADLNPNY 361
 Db 288 VTNQWVNSYKRIYVGNREAPTAATMGVNSRSALVRVPTVRLNKESSRREVRVLPDTACNPNY 347
 Qy 362 LAFSAIIAAGISGIEEKLPPPPASGNV-----YNDKELPEFPNSLQNAHLL 409
 Db 348 LAFSVMLGAGLKGIGKEGYELDEPAEDDISNLSFRERRAMGYND-----LPSSLDQALRQM 402
 Qy 410 KESKMLNKTGFEKILHYVNAANVEINEFQSKVQTDWELNQGDN 452
 Db 403 EKSELVADILGEHVFEEFLRNKREWRDYOQITPWEELRNLD 445
 RESULT 13
 Q8FNL7 PRELIMINARY; PRT; 516 AA.
 ID Q8FNL7
 AC Q8FNL7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Glutamine synthetase II (EC 6.3.1.2).
 DE GLNA2 OR CB2127.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.

Search completed: December 17, 2003, 22:44:19
Job time : 45 secs

